

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 18:05:34 ; Search time 44.971 Seconds
(without alignments)
102.356 Million cell updates/sec

Title: US-10-018-103A-4

Perfect score: 183

Sequence: 1 KHKKHKHKHKHKHKHKHKHKHKHK

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03.*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	183	100.0	29	AAE06229	Histidine copolyme
2	168.5	92.1	39	AAE06242	H-K2b polymer. Un
3	168.5	92.1	39	AAE06243	H-K3b polymer. Un
4	158	86.3	79	AAE06244	H-K4b polymer. Un
5	120	65.6	19	AAE06228	Histidine copolyme
6	120	65.6	21	AAE06234	Histidine copolyme
7	120	65.6	21	AAE06241	Histidine copolyme
8	117	63.9	29	AAE06232	Histidine copolyme
9	115	62.8	335	AAO17803	H Influenzae BVH-N

10	114	62.3	337	19	AAW61155	Haemophilus influe
11	110.5	60.4	1043	23	ABG96401	Human ovarian canc
12	104.5	57.1	271	22	ABG93454	Human protein sequ
13	104.5	57.1	504	22	ABG93135	Human protein sequ
14	104.5	57.1	1199	20	AAZ29083	Human CIP150/HTAFI
15	104.5	57.1	1199	21	AAZ44263	Human CIP150/HTAFI
16	98	53.6	129	22	AAU44422	Propionibacterium
17	96.5	52.7	62	22	AAE06246	HH-K3b polymer. U
18	96.5	52.7	83	22	AAE06247	HH-K4b polymer. U
19	96	52.5	980	23	ABG93037	S. cerevisiae BAX-
20	95.5	52.2	41	22	AAE06245	HH-K2b polymer. U
21	95	51.9	351	13	AAZ24393	Sequence of Histid
22	94	51.4	15	22	AAE06227	Histidine copolyme
23	93	50.8	337	22	ABG63629	Drosophila melanog
24	93	50.8	337	22	AAU38954	Drosophila G-prote
25	91	49.7	359	21	AAZ31823	Arabidopsis thalia
26	91	49.7	375	21	AAZ31822	Arabidopsis thalia
27	91	49.7	385	21	AAZ31821	Arabidopsis thalia
28	90	49.2	19	22	AAE06238	Histidine copolyme
29	89	48.6	81	22	AAO10602	Human polypeptide
30	88.5	48.4	639	22	ABG68805	Drosophila melanog
31	86	47.0	106	22	AAZ75565	Human colon cancer
32	86	47.0	483	22	AAZ92603	Human protein sequ
33	86	47.0	931	22	AAW40326	Human polypeptide
34	86	47.0	931	23	ABP65080	Hypoxia-repressed
35	84	45.9	78	20	AAZ07923	Human secreted pro
36	84	45.9	82	22	ABG47362	Human liver peptid
37	84	45.9	82	22	ABG27363	Human peptide #14
38	84	45.9	82	22	ABG32509	Peptide #15 encode
39	84	45.9	82	22	ABG18015	Protein #14 encode
40	84	45.9	82	22	AAZ53343	Human bone marrow
41	84	45.9	82	22	AAZ65721	Human brain expres
42	84	45.9	82	22	AAZ13580	Peptide #14 encode
43	84	45.9	82	22	AAZ25978	Peptide #15 encode
44	84	45.9	82	22	AAW03333	Peptide #15 encode
45	84	45.9	82	23	ABG35350	Human peptide enco

ALIGNMENTS

RESULT 1

AAE06229

ID AAE06229 standard; peptide; 29 AA.

XX

AC AAE06229;

XX

DT 25-SEP-2001 (first entry)

XX

DE Histidine copolymer, H-K 29-mer peptide.

XX

KW Histidine copolymer; enzyme deficiency; phenylalanine hydroxylase;
KW ornithine transcarbamylase; adenosine deaminase; tyrosine hydroxylase;
KW purine nucleoside phosphorylase; alpha-antitrypsin; apolipoprotein E;
KW chronic granulomatous disease; sickle cell; beta-thalassemia; anaemia;
KW familial hypercholesterolaemia; low-density lipoprotein receptor;
KW phenylketonuria; Faconi's anaemia; haemophilia; muscular dystrophy;
KW cystic fibrosis; Parkinson's disease; retinitis pigmentosa; cytostatic;
KW lysosomal storage disease; mycoplasma; type 1 disease; cardiant;
KW diabetic retinopathy; human immunodeficiency virus disease; infection;
KW cardiac disease; peripheral vascular disease; antisickling; arthritis;
KW neutropic; haemostatic; virucide; gene therapy.

XX Unidentified.

OS

XX WO200147496-A1.

PN

XX

PD 05-JUL-2001.

XX

XX 20-DEC-2000; 2000WO-US34603.

PF

XX

XX 29-DEC-1999; 99US-0173576.

PR

XX

KW lysosomal storage disease; mycopolysaccharide type 1 disease; cardiant;
 KW diabetic retinopathy; human immunodeficiency virus disease; infection;
 KW cardiac disease; peripheral vascular disease; antiskinking; arthritis;
 KW neutropic; haemostatic; virucide; gene therapy.
 OS Unidentified.
 XX
 PN WO200147496-A1.
 XX
 PD 05-JUL-2001.
 XX
 PF 20-DEC-2000; 2000WO-US34603.
 XX
 PR 29-DEC-1999; 99US-0173576.
 XX
 PA (MIXS/) MIXSON A J.
 XX
 PI Mixson AJ;
 XX
 DR WPI; 2001-425579/45.
 XX
 XX
 PT Pharmacutical composition useful for delivering therapeutic agent for
 PT treating diseases, comprises a peptide characterized with specified
 PT number of amino acids and a specified percentage of histidine residues
 PT .
 XX
 PS Example; Page 23; 64pp; English.
 XX
 CC The invention relates to a pharmaceutical agent delivery composition
 CC comprising a transport polymer comprising a linear or branched peptide
 CC having about 10 to 300 amino acid residues, having about 5 to 100%
 CC histidine residues, and optionally having 1 to 95% non-histidine
 CC residues, at least one pharmaceutical agent and optionally one or more
 CC intracellular delivery components in association with the transport
 CC polymer. The pharmaceutical agent delivery composition is useful for
 CC delivering a pharmaceutical agent to the interior of a cell and is
 CC used in gene therapy for treating adenosine deaminase deficiency, purine
 CC nucleoside phosphorylase deficiency, chronic granulomatous disease with
 CC defective p47phox, sickle cell with HbS, beta-thalassemia due to
 CC inadequate production of beta-haemoglobin, familial hypercholesterolaemia
 CC due to a defective low-density lipoprotein receptor, phenylketonuria due
 CC to phenylalanine hydroxylase deficiency, alpha1-antitrypsin deficiency,
 CC ornithine transcarbamylase deficiency, Faconi's anaemia, apolipoprotein
 CC E deficiency, haemophilia A and B due to factor VIII and IX deficiency,
 CC muscular dystrophy due to dystrophin, laminin-2, or sacroglycans
 CC mutations, cystic fibrosis due to CFTR mutations, Parkinson due to
 CC tyrosine hydroxylase deficiency, retinitis pigmentosa, lysosomal storage
 CC disease (i.e., mycopolysaccharide type 1, Hunter, Hurler and Gaucher),
 CC diabetic retinopathy, human immunodeficiency virus disease, virus
 CC infection, acquired anaemia, cardiac and peripheral vascular disease
 CC and arthritis. The present peptide sequence is H-K3b polymer.
 XX
 SQ Sequence 59 AA;
 Query Match 92.1%; Score 168.5; DB 22; Length 59;
 Best Local Similarity 76.3%; Pred. No. 8.1e-15;
 Matches 29; Conservative 0; Mismatches 0; Indels 9; Gaps 1;
 QY 1 KHGKHGKHGK-----KHGKHGKHGKHGKHGKHGKH 29
 DB 1 KHGKHGKHGKHGKHGKHGKHGKHGKHGKHGKHGKHGKH 38
 RESULT 4
 AAE06244
 ID AAE06244 standard; peptide; 79 AA.
 XX
 AC AAE06244;
 XX
 DT 25-SEP-2001 (first entry)
 XX
 DE H-K4b polymer.
 XX

KW Enzyme deficiency; phenylalanine hydroxylase;
 KW ornithine transcarbamylase; adenosine deaminase; tyrosine hydroxylase;
 KW purine nucleoside phosphorylase; alpha1-antitrypsin; apolipoprotein E;
 KW chronic granulomatous disease; sickle cell; beta-thalassemia; anaemia;
 KW familial hypercholesterolaemia; low-density lipoprotein receptor;
 KW phenylketonuria; Faconi's anaemia; haemophilia; muscular dystrophy;
 KW cystic fibrosis; Parkinson's disease; retinitis pigmentosa; cytostatic;
 KW lysosomal storage disease; mycopolysaccharide type 1 disease; cardiant;
 KW diabetic retinopathy; human immunodeficiency virus disease; infection;
 KW cardiac disease; peripheral vascular disease; antiskinking; arthritis;
 KW neutropic; haemostatic; virucide; gene therapy.
 XX
 OS Unidentified.
 XX
 PN WO200147496-A1.
 XX
 PD 05-JUL-2001.
 XX
 PF 20-DEC-2000; 2000WO-US34603.
 XX
 PR 29-DEC-1999; 99US-0173576.
 XX
 PA (MIXS/) MIXSON A J.
 XX
 PI Mixson AJ;
 XX
 DR WPI; 2001-425579/45.
 XX
 XX
 PT Pharmacutical composition useful for delivering therapeutic agent for
 PT treating diseases, comprises a peptide characterized with specified
 PT number of amino acids and a specified percentage of histidine residues
 PT .
 XX
 PS Example; Page 23; 64pp; English.
 XX
 CC The invention relates to a pharmaceutical agent delivery composition
 CC comprising a transport polymer comprising a linear or branched peptide
 CC having about 10 to 300 amino acid residues, having about 5 to 100%
 CC histidine residues, and optionally having 1 to 95% non-histidine
 CC residues, at least one pharmaceutical agent and optionally one or more
 CC intracellular delivery components in association with the transport
 CC polymer. The pharmaceutical agent delivery composition is useful for
 CC delivering a pharmaceutical agent to the interior of a cell and is
 CC used in gene therapy for treating adenosine deaminase deficiency, purine
 CC nucleoside phosphorylase deficiency, chronic granulomatous disease with
 CC defective p47phox, sickle cell with HbS, beta-thalassemia due to
 CC inadequate production of beta-haemoglobin, familial hypercholesterolaemia
 CC due to a defective low-density lipoprotein receptor, phenylketonuria due
 CC to phenylalanine hydroxylase deficiency, alpha1-antitrypsin deficiency,
 CC ornithine transcarbamylase deficiency, Faconi's anaemia, apolipoprotein
 CC E deficiency, haemophilia A and B due to factor VIII and IX deficiency,
 CC muscular dystrophy due to dystrophin, laminin-2, or sacroglycans
 CC mutations, cystic fibrosis due to CFTR mutations, Parkinson due to
 CC tyrosine hydroxylase deficiency, retinitis pigmentosa, lysosomal storage
 CC disease (i.e., mycopolysaccharide type 1, Hunter, Hurler and Gaucher),
 CC diabetic retinopathy, human immunodeficiency virus disease, virus
 CC infection, acquired anaemia, cardiac and peripheral vascular disease
 CC and arthritis. The present peptide sequence is H-K4b polymer.
 XX
 SQ Sequence 79 AA;
 Query Match 86.3%; Score 158; DB 22; Length 79;
 Best Local Similarity 74.4%; Pred. No. 2.3e-13;
 Matches 29; Conservative 0; Mismatches 0; Indels 10; Gaps 2;
 QY 1 KHGKHGKHGK-KGKHGKHGKH-----KGKHGKHGKHGKH 29
 DB 1 KHGKHGKHGKHGKHGKHGKHGKHGKHGKHGKHGKHGKH 39
 RESULT 5
 AAE06228
 ID AAE06228 standard; peptide; 19 AA.

CC K-HK.
 XX
 SQ Sequence 21 AA;
 Query Match 65.6%; Score 120; DB 22; Length 21;
 Best Local Similarity 100.0%; Pred. No. 4.1e-09;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KHKHKHKHKHKHKHKHK 19
 |||||
 DB 2 KHKHKHKHKHKHKHKHK 20
 |||||
 RESULT 7
 AAE06241
 ID AAE06241 standard; peptide; 21 AA.
 XX
 AC AAE06241;
 XX
 DT 25-SEP-2001 (first entry)
 XX
 DE Histidine copolymer, peptide #4.
 XX
 KW Histidine copolymer; enzyme deficiency; phenylalanine hydroxylase;
 KW ornithine transcarbamylase; adenosine deaminase; tyrosine hydroxylase;
 KW purine nucleoside phosphorylase; algal-antitrypsin; apolipoprotein E;
 KW chronic granulomatous disease; sickle cell; beta-thalassemia; anaemia;
 KW familial hypercholesterolaemia; low-density lipoprotein receptor;
 KW phenylketonuria; Faconi's anaemia; haemophilia; muscular dystrophy;
 KW cystic fibrosis; Parkinson's disease; retinitis pigmentosa; cytostatic;
 KW lysosomal storage disease; mycopolysaccharide type 1 disease; cardiant;
 KW diabetic retinopathy; human immunodeficiency virus disease; infection;
 KW cardiac disease; peripheral vascular disease; antisickling; arthritis;
 KW neutropic; haemostatic; virucide; gene therapy.
 XX
 OS Unidentified.
 XX
 PN WO200147496-A1.
 XX
 PD 05-JUL-2001.
 XX
 PF 20-DEC-2000; 2000WO-US34603.
 XX
 PR 29-DEC-1999; 99US-0173576.
 XX
 PA (MIXS/) MIXSON A J.
 XX
 PI Mixson AJ;
 XX
 DR WPI; 2001-425579/45.
 XX
 PT Pharmaceutical composition useful for delivering therapeutic agent for
 PT treating diseases, comprises a peptide characterized with specified
 PT number of amino acids and a specified percentage of histidine residues
 XX
 PS Claim 3; Page 36; 64pp; English.
 XX
 CC The invention relates to a pharmaceutical agent delivery composition
 CC comprising a transport polymer comprising a linear or branched peptide
 CC having about 10 to 300 amino acid residues, having about 5 to 100%
 CC histidine residues, and optionally having 1 to 95% non-histidine
 CC residues, at least one pharmaceutical agent and optionally one or more
 CC intracellular delivery components in association with the transport
 CC polymer. The pharmaceutical agent delivery composition is useful for
 CC delivering a pharmaceutical agent to the interior of a cell and is
 CC used in gene therapy for treating adenosine deaminase deficiency, purine
 CC nucleoside phosphorylase deficiency, chronic granulomatous disease with
 CC defective p47phox, sickle cell with HbS, beta-thalassemia due to
 CC inadequate production of beta-haemoglobin, familial hypercholesterolaemia
 CC due to a defective low-density lipoprotein receptor, phenylketonuria due
 CC to phenylalanine hydroxylase deficiency, algal-antitrypsin deficiency,
 CC ornithine transcarbamylase deficiency, Faconi's anaemia, apolipoprotein

CC E deficiency, haemophilia A and B due to factor VIII and IX deficiency,
 CC muscular dystrophy due to dystrophin, laminin-2, or sacroglucans
 CC mutations, cystic fibrosis due to CFTR mutations, Parkinson due to
 CC tyrosine hydroxylase deficiency, retinitis pigmentosa, lysosomal storage
 CC disease (i.e., mycopolysaccharide type 1, Hunter, Hurler and Gaucher),
 CC diabetic retinopathy, human immunodeficiency virus disease, virus
 CC infection, acquired anaemia, cardiac and peripheral vascular disease
 CC and arthritis. The present peptide sequence is a histidine copolymer.
 XX
 SQ Sequence 21 AA;
 Query Match 65.6%; Score 120; DB 22; Length 21;
 Best Local Similarity 100.0%; Pred. No. 4.1e-09;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KHKHKHKHKHKHKHKHK 19
 |||||
 DB 2 KHKHKHKHKHKHKHKHK 20
 |||||
 RESULT 8
 AAE06232
 ID AAE06232 standard; peptide; 29 AA.
 XX
 AC AAE06232;
 XX
 DT 25-SEP-2001 (first entry)
 XX
 DE Histidine copolymer, Y-HK peptide.
 XX
 KW Histidine copolymer; enzyme deficiency; phenylalanine hydroxylase;
 KW ornithine transcarbamylase; adenosine deaminase; tyrosine hydroxylase;
 KW purine nucleoside phosphorylase; algal-antitrypsin; apolipoprotein E;
 KW chronic granulomatous disease; sickle cell; beta-thalassemia; anaemia;
 KW familial hypercholesterolaemia; low-density lipoprotein receptor;
 KW phenylketonuria; Faconi's anaemia; haemophilia; muscular dystrophy;
 KW cystic fibrosis; Parkinson's disease; retinitis pigmentosa; cytostatic;
 KW lysosomal storage disease; mycopolysaccharide type 1 disease; cardiant;
 KW diabetic retinopathy; human immunodeficiency virus disease; infection;
 KW cardiac disease; peripheral vascular disease; antisickling; arthritis;
 KW neutropic; haemostatic; virucide; gene therapy.
 XX
 OS Unidentified.
 XX
 FH Key Location/Qualifiers
 FT Region 1..11
 FT /note= "Y component"
 XX
 PN WO200147496-A1.
 XX
 PD 05-JUL-2001.
 XX
 PF 20-DEC-2000; 2000WO-US34603.
 XX
 PR 29-DEC-1999; 99US-0173576.
 XX
 PA (MIXS/) MIXSON A J.
 XX
 PI Mixson AJ;
 XX
 DR WPI; 2001-425579/45.
 XX
 PT Pharmaceutical composition useful for delivering therapeutic agent for
 PT treating diseases, comprises a peptide characterized with specified
 PT number of amino acids and a specified percentage of histidine residues
 XX
 PS Example; Page 23; 64pp; English.
 XX
 CC The invention relates to a pharmaceutical agent delivery composition
 CC comprising a transport polymer comprising a linear or branched peptide
 CC having about 10 to 300 amino acid residues, having about 5 to 100%
 CC histidine residues, and optionally having 1 to 95% non-histidine

PR 25-JUL-1996; 96CA-2182046.
XX
PA (HSCR-) HSC RES & DEV LP.
XX
PI Lingwood CA;
v

cells or from transformed host cells). Adhesin polyepitopes of nucleic acid molecules (see AAV36458) encoding them can also be used in claimed immunogenic compositions formulated as vaccines to protect a mammal (preferably a human) against diseases caused by bacterial pathogens that have the adhesin as a surface protein.

SQ	Sequence	337 AA;
	Query Match	62.3%; Score 114; DB 19; Length 337;
	Best Local Similarity	60.7%; Pred. No. 3.9e-07;
	Conservation	2. Mismatches 8; Indels 0; Gaps

[illegible]

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification, the primer sets can be used in antisense therapy and

Job time : 45.971 secs

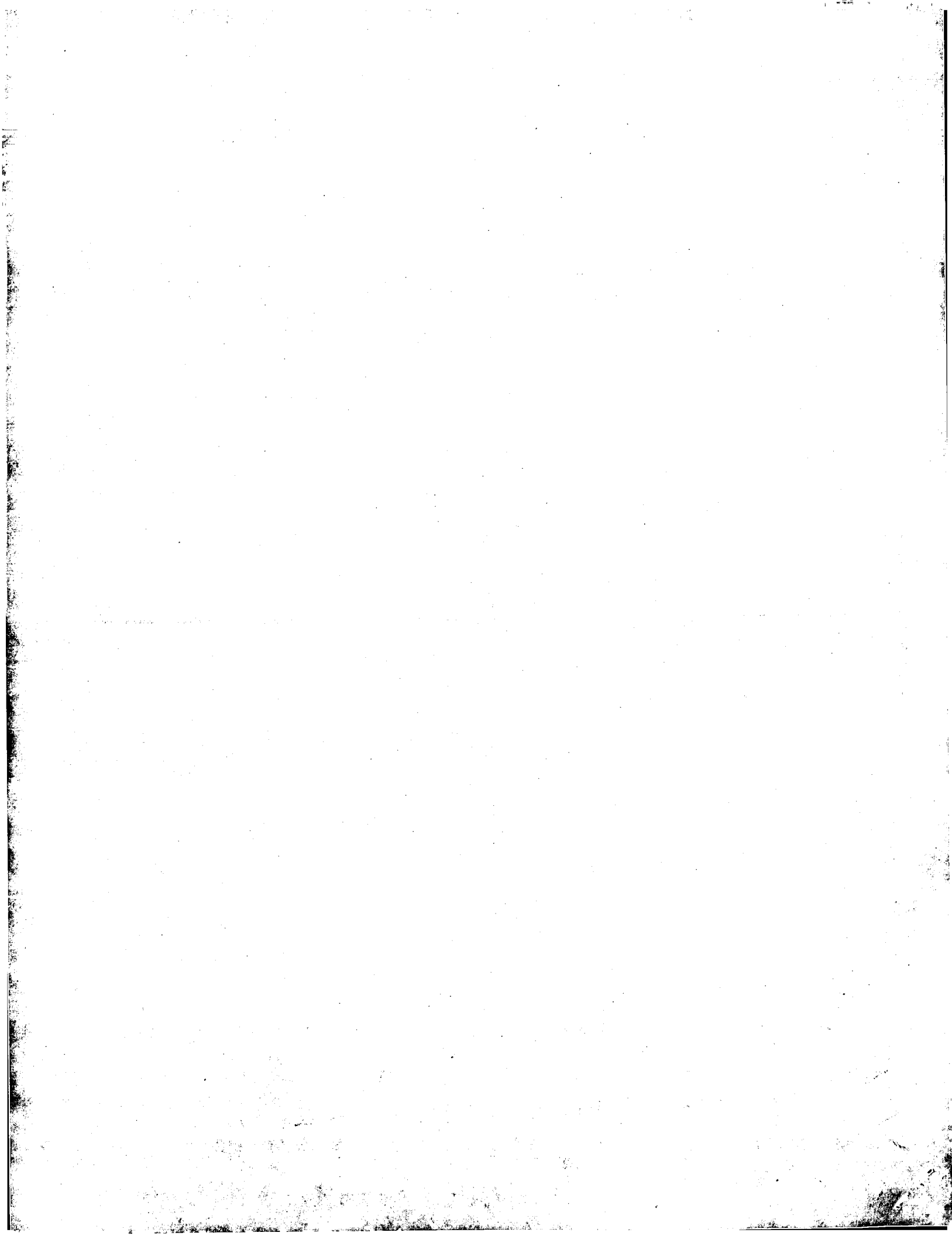
CC either in vitro, such as in a cell culture, or in vivo. The products can
CC also be used for the diagnosis, prognosis and predicting susceptibility
CC to neoplasia.

SQ Sequence 1199 AA;
Query Match 57.1%; Score 104.5; DB 20; Length 1199;
Best Local Similarity 67.9%; Pred. No. 2.3e-05;
Matches 19; Conservative 1; Mismatches 7; Indels 1; Gaps 1;
Qy 2 HKKKHKHKHKHKHKHKHKHKHKHK 29
Db 1148 HHHHEHK-KKKKKHKHKHKHKHDSK 1174

RESULT 15
AA44263
ID AAY44263 standard; Protein; 1199 AA.
XX
AC AAY44263;
XX
DT 22-MAR-2000 (first entry)
XX
DE Human C1F150/htAF1150 essential for mitosis.
XX
KW Human; C1F130 protein; cell cycle regulator; mitosis; neoplasia;
KW C1F150/htAF1150; diagnosis; treatment; proliferative disorder;
KW hyperplasia; dysplasia; C1F150.
XX
OS Homo sapiens.
XX
PN WO9964450-A2.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US13143.
XX
PR 12-JUN-1998; 98US-0089198.
PR 09-DEC-1998; 98US-0111636.
XX
PA (CHIR) CHIRON CORP.
XX
PI Kaufmann J;
XX
DR WPI: 2000-116524/10.
DR N-PSDB; AA229605.
XX
PT Novel regulatory protein, useful to diagnose and treat cell
PT proliferation disorders, including cancer -
XX
PS Claim 20; Fig 3; 52pp; English.
XX
CC The present sequence is a human C1F150/htAF1150 protein
CC which is required for a human cell to enter mitosis. In particular,
CC C1F150/htAF1150 is an essential cofactor for TPID-dependent
CC transcription. C1F130 negatively regulates C1F150 thereby inhibiting
CC mitosis or cell cycle progression. C1F150 has a C1F130-binding
CC domain and is used to screen compounds that interfere with the binding
CC of C1F130. C1F130 is used to diagnose and treat proliferative disorders
CC including hyperplasias, neoplasias and dysplasias.

SQ Sequence 1199 AA;
Query Match 57.1%; Score 104.5; DB 21; Length 1199;
Best Local Similarity 67.9%; Pred. No. 2.3e-05;
Matches 19; Conservative 1; Mismatches 7; Indels 1; Gaps 1;
Qy 2 HKKKHKHKHKHKHKHKHKHKHKHK 29
Db 1148 HHHHEHK-KKKKKHKHKHKHKHDSK 1174

Search completed: January 20, 2004, 18:23:18



GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 18:22:04 ; Search time 16.3913 Seconds
(without alignments)
74.858 Million cell updates/sec

Title: US-10-018-103A-4

Perfect score: 183

Sequence: 1 KHKKHKHKHKHKHKHKHKHKHKHKHK

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	114	62.3	313	3	US-08-686-528A-3
2	114	62.3	313	3	US-09-456-287-3
3	114	62.3	337	3	US-08-686-528A-2
4	114	62.3	337	3	US-09-456-287-2
5	104.5	57.1	1199	3	US-09-208-742-2
6	104.5	57.1	1199	4	US-09-332-295-4
7	104.5	57.1	1199	4	US-09-709-979-4
8	84	45.9	726	3	US-09-126-980-2
9	84	45.9	726	3	US-09-476-482-2
10	84	45.9	726	4	US-09-517-605-6
11	83.5	45.6	300	4	US-09-395-689-1
12	83.5	45.6	765	2	US-08-663-112-2
13	80	43.7	94	3	US-09-612-126-10
14	80	43.7	179	3	US-09-126-126-11
15	80	43.7	186	3	US-09-612-126-8
16	80	43.7	255	3	US-09-612-126-1
17	79.5	43.4	47	3	US-09-612-126-4
18	79.5	43.4	62	3	US-09-612-126-7
19	79.5	43.4	83	3	US-09-612-126-6
20	78	42.6	18	1	US-08-346-849-64
21	78	42.6	18	2	US-08-293-284A-64
22	78	42.6	18	4	US-08-898-300-64
23	78	42.6	344	4	US-09-134-001C-3524
24	77	42.1	150	4	US-09-395-689-2
25	76.5	41.8	533	4	US-09-252-991A-23560
26	76	41.5	261	4	US-09-602-565-34
27	75	41.0	218	4	US-09-252-991A-25291

28	74.5	40.7	1213	1	US-08-188-582-20	Sequence 20, Appl
29	74.5	40.7	1213	1	US-08-646-715-20	Sequence 20, Appl
30	72.5	39.6	1664	1	US-09-599-652-2	Sequence 2, Appli
31	72.5	39.6	1664	2	US-08-642-846-2	Sequence 2, Appli
32	72.5	39.6	1664	4	US-09-264-604-2	Sequence 2, Appli
33	72	39.3	363	4	US-09-328-352-4930	Sequence 4930, Ap
34	71	38.8	40	4	US-09-039-780A-6	Sequence 6, Appli
35	71	38.8	83	4	US-09-420-592A-10	Sequence 10, Appl
36	70.5	38.5	474	3	US-09-461-474-10	Sequence 10, Appl
37	70	38.3	349	3	US-09-461-474-12	Sequence 12, Appl
38	69.5	38.0	316	4	US-09-252-991A-27084	Sequence 27084, A
39	65	35.5	203	4	US-09-252-991A-26395	Sequence 26395, A
40	64.5	35.2	60	1	US-08-255-457-1	Sequence 1, Appli
41	64.5	35.2	60	2	US-09-115-032-1	Sequence 1, Appli
42	64.5	35.2	60	5	PCT-US95-05772-1	Sequence 1, Appli
43	64	35.0	148	4	US-09-461-325-453	Sequence 453, App
44	64	35.0	298	3	US-09-461-474-17	Sequence 17, Appl
45	63	34.4	244	4	US-09-252-991A-30724	Sequence 30724, A

ALIGNMENTS

RESULT 1
US-08-686-528A-3
; Sequence 3, Application US/08686528A
; Patent No. 6054134
; GENERAL INFORMATION:
; APPLICANT: LINGWOOD, Clifford A.
; TITLE OF INVENTION: HAEMOPHILUS ADHESIN PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Burns, Doane, Swecker & Mathis, L.L.P.
; STREET: 1737 King Street, Suite 500
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22314-2756
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/686,528A
; FILING DATE: 26-JUL-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rea, Teresa Stanek
; REGISTRATION NUMBER: 30,427
; REFERENCE/DOCKET NUMBER: 032609-001
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 313 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-686-528A-3

Query Match 62.3%; Score 114; DB 3; Length 313;
Best Local Similarity 60.7%; Pred. No. 4.7e-07;
Matches 17; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 2 HKHKHKHKHKHKHKHKHKHKHKHK 29
DB 105 HKHKHKHKHKHKHKHKHKHKHK 132

RESULT 2
US-09-456-287-3
; Sequence 3, Application US/09456287
; Patent No. 6218147

Wed Jan 21 11:27:48 2004

```
;
;
; GENERAL INFORMATION:
; APPLICANT: LINGWOOD, Clifford A.
; TITLE OF INVENTION: HAEMOPHILUS ADHESIN PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis, L.L.P.
; STREET: 1737 King Street, Suite 500
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22314-2756
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/456,287
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: US 08/686,528
; FILING DATE:
; APPLICANT INFORMATION:
; NAME: Rea, Teresa Stanek
; REGISTRATION NUMBER: 30,427
; REFERENCE/DOCKET NUMBER: 032609-001
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 313 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-456-287-3
;
; Query Match 62.3%; Score 114; DB 3; Length 313;
; Best Local Similarity 60.7%; Pred. No. 4.7e-07;
; Matches 17; Conservative 3; Mismatches 8; Indels 0; Gaps 0;
;
; QY 2 HKHKHKHKHKHKHKHKHKHKHK 29
; DB 105 HKHEKHDKHDKHDKHDKHDKHDKHE 132
;
; RESULT 3
; US-08-686-528A-2
; Sequence 2, Application US/08686528A
; Patent No. 6054134
; GENERAL INFORMATION:
; APPLICANT: LINGWOOD, Clifford A.
; TITLE OF INVENTION: HAEMOPHILUS ADHESIN PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis, L.L.P.
; STREET: 1737 King Street, Suite 500
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22314-2756
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 26-JUL-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rea, Teresa Stanek
; REGISTRATION NUMBER: 30,427
; REFERENCE/DOCKET NUMBER: 032609-001
;
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 337 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-456-287-2
;
; Query Match 62.3%; Score 114; DB 3; Length 337;
; Best Local Similarity 60.7%; Pred. No. 5.1e-07;
; Matches 17; Conservative 3; Mismatches 8; Indels 0; Gaps 0;
;
; QY 2 HKHKHKHKHKHKHKHKHKHKHK 29
; DB 129 HKHEKHDKHDKHDKHDKHDKHDKHE 156
;
; RESULT 4
; US-09-456-287-2
; Sequence 2, Application US/09456287
; Patent No. 6218147
; GENERAL INFORMATION:
; APPLICANT: LINGWOOD, Clifford A.
; TITLE OF INVENTION: HAEMOPHILUS ADHESIN PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis, L.L.P.
; STREET: 1737 King Street, Suite 500
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22314-2756
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: US 08/686,528
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Rea, Teresa Stanek
; REGISTRATION NUMBER: 30,427
; REFERENCE/DOCKET NUMBER: 032609-001
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 337 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-456-287-2
;
; Query Match 62.3%; Score 114; DB 3; Length 337;
; Best Local Similarity 60.7%; Pred. No. 5.1e-07;
; Matches 17; Conservative 3; Mismatches 8; Indels 0; Gaps 0;
;
; QY 2 HKHKHKHKHKHKHKHKHKHKHK 29
; DB 129 HKHEKHDKHDKHDKHDKHDKHDKHE 156
;
; RESULT 5
; US-09-208-742-2
; Sequence 2, Application US/09208742
; Patent No. 6174679
; GENERAL INFORMATION:
; APPLICANT: Kaufmann, Joerg
```



```

1  TITLE OF INVENTION: MUTANT PROTEINS OF HUMAN DNA
2  TITLE OF INVENTION: TOPOISOMERASE I
3  NUMBER OF SEQUENCES: 7
4  CORRESPONDENCE ADDRESS:
5  ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
6  ADDRESSEE: Dunner L.L.P.
7  STREET: 1300 I Street, N.W.
8  CITY: Washington
9  STATE: D.C.
10 COUNTRY: USA
11 ZIP: 20005-3315
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE: Floppy disk
14 COMPUTER: IBM PC compatible
15 OPERATING SYSTEM: PC-DOS/MS-DOS
16 SOFTWARE: PatentIn Release #1.0, Version #1.30
17 CURRENT APPLICATION DATA:
18 APPLICATION NUMBER: US/08/663,112
19 FILING DATE: 26-NOV-1996
20 CLASSIFICATION: 435
21 ATTORNEY/AGENT INFORMATION:
22 NAME: Einaudi, Carolyn P.
23 REGISTRATION NUMBER: 32,220
24 REFERENCE/DOCKET NUMBER: 06609,1488-00000
25 TELECOMMUNICATION INFORMATION:
26 TELEPHONE: 202-408-4000
27 TELEFAX: 202-408-4400
28 INFORMATION FOR SEQ ID NO: 2:
29 SEQUENCE CHARACTERISTICS:
30 LENGTH: 765 amino acids
31 TYPE: amino acid
32 TOPOLOGY: linear
33 MOLECULE TYPE: protein
34 US-08-663-112-2
35
36 Query Match . 45.6%; Score 83.5; DB 2; Length 765;
37 Best Local Similarity 36.1%; Pred. No. 0.0047; Indels 33; Gaps 3
38 Matches 22; Conservative 0; Mismatches 6;
39
40 QY 2 HKHKHKH-----KGKH-----KHKHKHKGKH-----KHKHKH 28
41 Db 22 HKHKHKHDKRHHKHKHKDKRKSNSNKHDKSEKHKHKEKTKHKDGSSEKHKDKH 81
42 QY 29 K 29
43 Db 82 K 82
44
45 RESULT 13
46 US-09-612-126-10
47 Sequence 10, Application US/09612126
48 Patent No. 6284726
49 GENERAL INFORMATION:
50 APPLICANT: TEMPLE UNIVERSITY - OF THE COMMONWEALTH SYSTEM OF
51 TITLE OF INVENTION: INHIBITION OF ANGIOGENESIS BY PEPTIDE ANALOGS OF HIGH
52 FILE REFERENCE: MOLECULAR WEIGHT KININOGEN DOMAIN 5
53 CURRENT APPLICATION NUMBER: US/09/612,126
54 CURRENT FILING DATE: 2000-07-07
55 PRIOR APPLICATION NUMBER: 60/107,844
56 PRIOR FILING DATE: 1998-11-10
57 PRIOR APPLICATION NUMBER: PCT/US99/26377
58 PRIOR FILING DATE: 1999-11-09
59 NUMBER OF SEQ ID NOS: 12
60 SOFTWARE: PatentIn Ver. 2.1
61 SEQ ID NO 10
62 LENGTH: 94
63 TYPE: PRT
64 ORGANISM: Artificial Sequence
65 FEATURE:
66 OTHER INFORMATION: Description of Artificial Sequence: Human high
67 OTHER INFORMATION: molecular weight kininogen light chain amino acids
68 OTHER INFORMATION: Lys(420) through Ser(513)

```

```

; LABEL: FR1
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Human high
; OTHER INFORMATION: molecular weight kininogen light chain amino acids
; OTHER INFORMATION: His(441) through Ser(626)
US-09-612-126-8

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 18:23:25 ; Search time 32.7826 Seconds
(without alignments)
180.887 Million cell updates/sec

Title: US-10-018-103A-4

Perfect score: 183

Sequence: 1 KHKHKHKHKHKHKHKHKHKHKHK

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 762491 seqs, 204481190 residues

Total number of hits satisfying chosen parameters: 762491

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pap.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pap.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pap.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pap.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pap.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pap.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pap.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pap.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pap.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pap.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pap.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pap.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pap.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pap.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pap.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pap.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pap.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	183	100.0	29	12	US-10-131-909A-4
2	183	100.0	29	15	US-10-018-103A-4
3	120	65.6	19	12	US-10-131-909A-3
4	120	65.6	19	15	US-10-018-103A-3
5	120	65.6	21	12	US-10-131-909A-9
6	120	65.6	21	12	US-10-131-909A-16
7	120	65.6	21	15	US-10-018-103A-9
8	120	65.6	21	15	US-10-018-103A-16
9	117	63.9	29	12	US-10-131-909A-7
10	117	63.9	29	15	US-10-018-103A-7
11	110.5	60.4	1007	15	US-10-211-133-7
12	110.5	60.4	1043	15	US-10-097-340-258
13	104.5	57.1	1199	14	US-10-147-268-4
14	104.5	57.1	1199	15	US-10-338-279-4
15	96	52.5	980	12	US-10-369-493-1406

16	94	51.4	15	12	US-10-131-909A-2	Sequence 2, Appli
17	94	51.4	15	15	US-10-018-103A-2	Sequence 2, Appli
18	93	50.8	337	15	US-10-270-333-96	Sequence 96, Appli
19	91	49.7	574	15	US-10-156-761-14106	Sequence 14106, A
20	90	49.2	19	12	US-10-131-909A-13	Sequence 13, Appli
21	90	49.2	19	15	US-10-018-103A-13	Sequence 13, Appli
22	86	47.0	106	15	US-10-106-698-6339	Sequence 6339, Ap
23	86	47.0	885	12	US-10-108-260A-3459	Sequence 3459, Ap
24	85	46.4	378	12	US-10-029-386-33892	Sequence 33892, A
25	84	45.9	78	12	US-10-195-730-186	Sequence 186, App
26	84	45.9	82	9	US-09-864-761-33313	Sequence 33313, A
27	84	45.9	726	11	US-09-932-257A-19	Sequence 19, Appli
28	84	45.9	1257	12	US-10-369-493-6761	Sequence 6761, Ap
29	83.5	45.6	765	12	US-09-882-274-2	Sequence 2, Appli
30	82	44.8	49	9	US-09-864-761-37882	Sequence 37882, A
31	81.5	44.5	20	12	US-10-131-909A-5	Sequence 5, Appli
32	81.5	44.5	20	15	US-10-018-103A-5	Sequence 5, Appli
33	81	44.3	13	12	US-10-131-909A-1	Sequence 1, Appli
34	81	44.3	13	15	US-10-018-103A-1	Sequence 1, Appli
35	81	44.3	444	12	US-10-360-849A-30	Sequence 30, Appli
36	80	43.7	148	9	US-09-908-711-91	Sequence 91, Appli
37	80	43.7	148	11	US-09-764-891-3203	Sequence 3203, Ap
38	79.5	43.4	20	12	US-10-131-909A-6	Sequence 6, Appli
39	79.5	43.4	20	15	US-10-018-103A-6	Sequence 6, Appli
40	79	43.2	39	12	US-10-105-232-393	Sequence 393, App
41	79	43.2	39	12	US-10-189-437-380	Sequence 380, App
42	79	43.2	90	12	US-10-315-515-39	Sequence 39, Appli
43	79	43.2	90	12	US-10-315-515-44	Sequence 44, Appli
44	79	43.2	93	12	US-10-315-515-46	Sequence 46, Appli
45	79	43.2	95	12	US-10-315-515-35	Sequence 35, Appli

ALIGNMENTS

RESULT 1
US-10-131-909A-4
; Sequence 4, Application US/10131909A
; Publication No. US20030165567A1
; GENERAL INFORMATION:
; APPLICANT: Mixson, A. James
; TITLE OF INVENTION: Histidine Copolymer and Methods For Using Same
; FILE REFERENCE: 5627*6
; CURRENT APPLICATION NUMBER: US/10/131,909A
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US 10/018103
; PRIOR FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: PCT/US00/34603
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/173576
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 4
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
US-10-131-909A-4

Query Match 100.0%; Score 183; DB 12; Length 29;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KHKHKHKHKHKHKHKHKHKHKHK 29
Db 1 KHKHKHKHKHKHKHKHKHKHKHK 29

RESULT 2
US-10-018-103A-4
; Sequence 4, Application US/10018103A

```
; Publication No. US20030045465A1
; GENERAL INFORMATION:
; APPLICANT: Mixson, A. James
; TITLE OF INVENTION: Histidine Copolymer and Methods For Using Same
; FILE REFERENCE: 5627*5
; CURRENT APPLICATION NUMBER: US/10/018,103A
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US 60/173576
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: PCT/US00/34603
; PRIOR FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
US-10-018-103A-4

Query Match      100.0%; Score 183; DB 15; Length 29;
Best Local Similarity 100.0%; Pred. No. 1.4e-13;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KHKHKHKHKHKHKHKHKHKHKHKHK 29
Db 1 KHKHKHKHKHKHKHKHKHKHKHKHK 29

RESULT 3
US-10-131-909A-3
; Sequence 3, Application US/10131909A
; Publication No. US20030165567A1
; GENERAL INFORMATION:
; APPLICANT: Mixson, A. James
; TITLE OF INVENTION: Histidine Copolymer and Methods For Using Same
; FILE REFERENCE: 5627*6
; CURRENT APPLICATION NUMBER: US/10/131,909A
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US 10/018103
; PRIOR FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: PCT/US00/34603
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/173576
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
US-10-131-909A-3

Query Match      65.6%; Score 120; DB 12; Length 19;
Best Local Similarity 100.0%; Pred. No. 5.2e-07;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KHKHKHKHKHKHKHKHKHKHKHK 19
Db 1 KHKHKHKHKHKHKHKHKHKHKHK 20

RESULT 4
US-10-018-103A-3
; Sequence 3, Application US/10018103A
; Publication No. US20030045465A1
; GENERAL INFORMATION:
; APPLICANT: Mixson, A. James
; TITLE OF INVENTION: Histidine Copolymer and Methods For Using Same
; FILE REFERENCE: 5627*5
```

```
; CURRENT APPLICATION NUMBER: US/10/018,103A
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US 60/173576
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: PCT/US00/34603
; PRIOR FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
US-10-018-103A-3

Query Match      65.6%; Score 120; DB 15; Length 19;
Best Local Similarity 100.0%; Pred. No. 5.2e-07;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KHKHKHKHKHKHKHKHKHKHKHK 19
Db 1 KHKHKHKHKHKHKHKHKHKHKHK 19

RESULT 5
US-10-131-909A-9
; Sequence 9, Application US/10131909A
; Publication No. US20030165567A1
; GENERAL INFORMATION:
; APPLICANT: Mixson, A. James
; TITLE OF INVENTION: Histidine Copolymer and Methods For Using Same
; FILE REFERENCE: 5627*6
; CURRENT APPLICATION NUMBER: US/10/131,909A
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US 10/018103
; PRIOR FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: PCT/US00/34603
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/173576
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
US-10-131-909A-9

Query Match      65.6%; Score 120; DB 12; Length 21;
Best Local Similarity 100.0%; Pred. No. 5.7e-07;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KHKHKHKHKHKHKHKHKHKHKHK 19
Db 2 KHKHKHKHKHKHKHKHKHKHKHK 20

RESULT 6
US-10-131-909A-16
; Sequence 16, Application US/10131909A
; Publication No. US20030165567A1
; GENERAL INFORMATION:
; APPLICANT: Mixson, A. James
; TITLE OF INVENTION: Histidine Copolymer and Methods For Using Same
; FILE REFERENCE: 5627*6
; CURRENT APPLICATION NUMBER: US/10/131,909A
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US 10/018103
; PRIOR FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: PCT/US00/34603
```

; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/173576
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
US-10-131-909A-16

Query Match 65.6%; Score 120; DB 12; Length 21;
Best Local Similarity 100.0%; Pred. No. 5.7e-07;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KHKHKHKHKHKHKHKHK 19
Db 2 KHKHKHKHKHKHKHKHK 20

RESULT 7
US-10-018-103A-9
; Sequence 9, Application US/10018103A
; Publication No. US20030045465A1
; GENERAL INFORMATION:
; APPLICANT: Mixson, A. James
; TITLE OF INVENTION: Histidine Copolymer and Methods For Using Same
; FILE REFERENCE: 5627*5
; CURRENT APPLICATION NUMBER: US/10/018,103A
; PRIOR FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US 60/173576
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: PCT/US00/34603
; PRIOR FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
US-10-018-103A-9

Query Match 65.6%; Score 120; DB 15; Length 21;
Best Local Similarity 100.0%; Pred. No. 5.7e-07;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KHKHKHKHKHKHKHKHK 19
Db 2 KHKHKHKHKHKHKHKHK 20

RESULT 8
US-10-018-103A-16
; Sequence 16, Application US/10018103A
; Publication No. US20030045465A1
; GENERAL INFORMATION:
; APPLICANT: Mixson, A. James
; TITLE OF INVENTION: Histidine Copolymer and Methods For Using Same
; FILE REFERENCE: 5627*5
; CURRENT APPLICATION NUMBER: US/10/018,103A
; PRIOR FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US 60/173576
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: PCT/US00/34603
; PRIOR FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 21

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
US-10-018-103A-16

Query Match 65.6%; Score 120; DB 15; Length 21;
Best Local Similarity 100.0%; Pred. No. 5.7e-07;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KHKHKHKHKHKHKHKHK 19
Db 2 KHKHKHKHKHKHKHKHK 20

RESULT 9
US-10-131-909A-7
; Sequence 7, Application US/10131909A
; Publication No. US20030165567A1
; GENERAL INFORMATION:
; APPLICANT: Mixson, A. James
; TITLE OF INVENTION: Histidine Copolymer and Methods For Using Same
; FILE REFERENCE: 5627*6
; CURRENT APPLICATION NUMBER: US/10/131,909A
; PRIOR FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US 10/018103
; PRIOR FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: PCT/US00/34603
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/173576
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
US-10-131-909A-7

Query Match 63.9%; Score 117; DB 12; Length 29;
Best Local Similarity 66.7%; Pred. No. 1.6e-06;
Matches 18; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 1 KHKHKHKHKHKHKHKHK 27
Db 3 RKRRQRKHKHKHKHKHKHKHK 29

RESULT 10
US-10-018-103A-7
; Sequence 7, Application US/10018103A
; Publication No. US20030045465A1
; GENERAL INFORMATION:
; APPLICANT: Mixson, A. James
; TITLE OF INVENTION: Histidine Copolymer and Methods For Using Same
; FILE REFERENCE: 5627*5
; CURRENT APPLICATION NUMBER: US/10/018,103A
; PRIOR FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US 60/173576
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: PCT/US00/34603
; PRIOR FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
US-10-018-103A-7

```

; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 258
; LENGTH: 1043
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-340-258

Query Match          60.4%; Score 110.5; DB 15; Length 1043;
Best Local Similarity 59.0%; Pred. No. 0.00018;
Matches 23; Conservative 1; Mismatches 4; Indels 11; Gaps 3;

QY      1 KH-----KHKKHKGKHKHKKHGK-----KHKKHKKHKK 29
           |||:|||||||
Db       75 KHSRHHKKKHRSRKH-KHGSSEEDKOKKKHKKHKK 112

RESULT 13
US-10-147-268-4
; Sequence 4, Application US/10147268
; Publication No. US20020143154A1
; GENERAL INFORMATION:
; APPLICANT: Kaufmann, Joerg
; TITLE OF INVENTION: CIFI30 INHIBITS CELL CYCLE PROGRESSION
; FILE REFERENCE: PP-01513.104/200130.456D3
; CURRENT APPLICATION NUMBER: US/10/147,268
; CURRENT FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1199
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-147-268-4

Query Match          57.1%; Score 104.5; DB 14; Length 1199;
Best Local Similarity 67.9%; Pred. No. 0.0009;
Matches 19; Conservative 1; Mismatches 7; Indels 1; Gaps 1;

QY      2 HXKHGKHGKHKKHKKHKKHKKHKKHKK 29
           |||:|||||||
Db       1148 HHHHEHK-KKKKKHKKHKKHGGDSK 1174

RESULT 14
US-10-338-279-4
; Sequence 4, Application US/10138279
; Publication No. US20030113791A1
; GENERAL INFORMATION:
; APPLICANT: Kaufmann, Joerg
; TITLE OF INVENTION: CIFI30 INHIBITS CELL CYCLE PROGRESSION
; FILE REFERENCE: PP-01513.104/200130.456D3
; CURRENT APPLICATION NUMBER: US/10/338,279
; CURRENT FILING DATE: 2003-01-07
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1199
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-338-279-4

Query Match          57.1%; Score 104.5; DB 15; Length 1199;
Best Local Similarity 67.9%; Pred. No. 0.0009;
Matches 19; Conservative 1; Mismatches 7; Indels 1; Gaps 1;

```

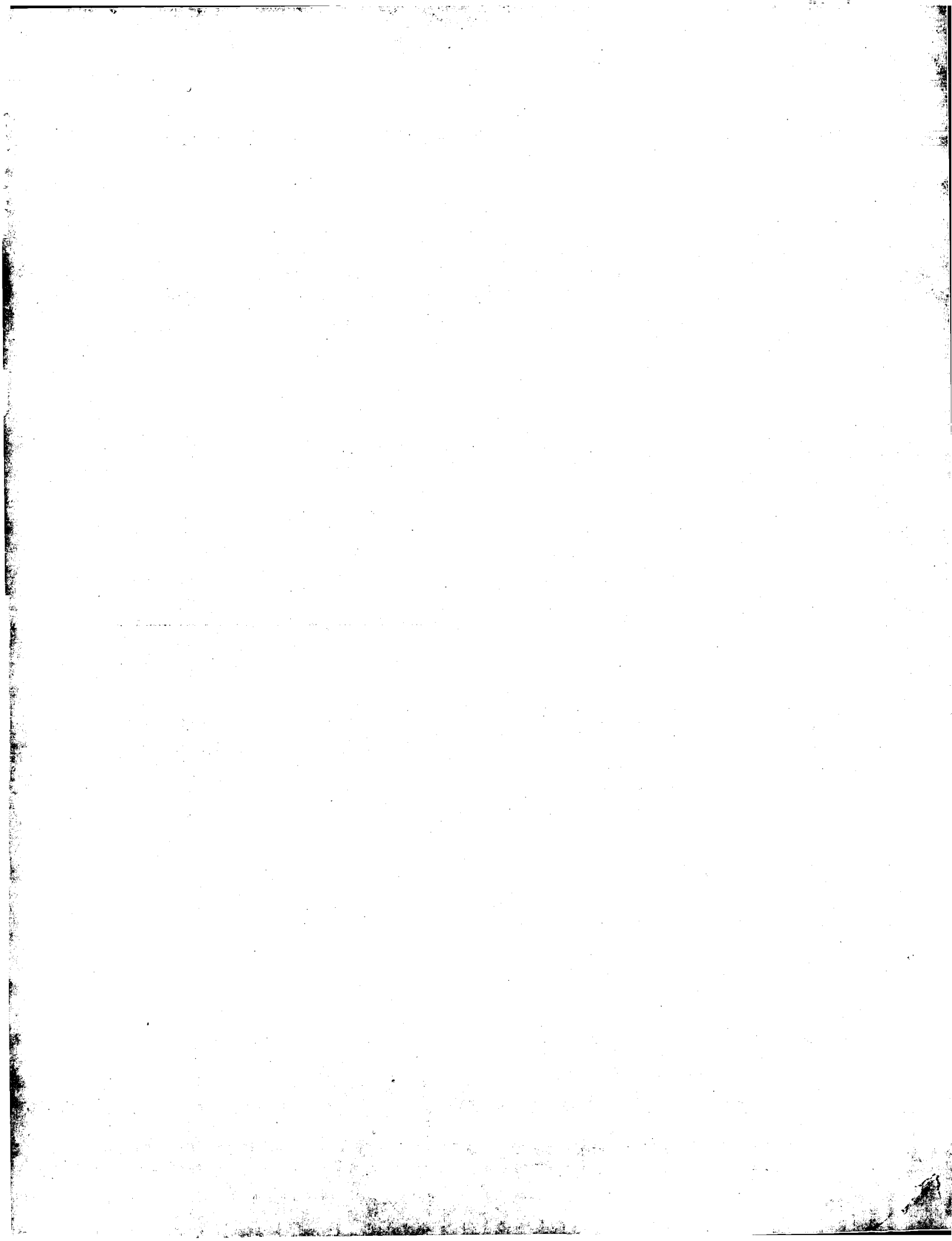

Qy 2 HKHKHKHKHKHKHKHKHKHKHKHK 29
| | | | | | | | | | | | | | | |
Db 1148 HHHHEHK-KKKKKHKHKHKHKHDSK 1174

RESULT 15
US-10-369-493-1406
; Sequence 1406, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1406
; LENGTH: 980
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-1406

Query Match 52.5%; Score 96; DB 12; Length 980;
Best Local Similarity 83.3%; Pred. No. 0.0061;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 9 KGKHKHKHKHKHKHKHKHKHKHKHK 26
: | | | | | | | | | | | | | | | |
Db 685 EGKHKHKHKHKHKHKHKHKHKKN 702

Search completed: January 20, 2004, 18:28:42
Job time : 32.7826 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 18:18:04 ; Search time 14.7101 Seconds
(without alignments)
189.590 Million cell updates/sec

Title: US-10-018-103A-4

Perfect score: 183

Sequence: 1 KHKKHKHKHKHKHKHKHKHKHKHK 29

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.76.*

2: PIR.*

3: PIR.*

4: PIR.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	114	62.3	337	D64049	adhesin homolog HI
2	96	52.5	980	S45444	BEH1 protein-bindi
3	95	51.9	351	KGZQHL	histidine-rich gly
4	93	50.8	496	S33791	ARS-binding protei
5	91	49.7	385	A84696	probable zinc tran
6	86	47.0	668	A44863	trophozoite antige
7	85	46.4	102	T07078	cold stress protei
8	84	45.9	283	C85838	hypothetical prote
9	84	45.9	1257	T28937	hypothetical prote
10	83.5	45.6	765	1TSHUT1	DNA topoisomerase
11	83	45.4	140	A54523	hypothetical-rich pro
12	82.5	45.1	744	T13429	hypothetical prote
13	82.5	45.1	767	JU0144	DNA topoisomerase
14	82.5	45.1	767	A49546	DNA topoisomerase
15	81.5	44.5	535	S66148	gene pipequeak pro
16	81.5	44.5	1085	S66149	gene pipequeak pro
17	81.5	44.5	1388	T00063	hypothetical prote
18	81	44.3	658	T04219	hypothetical prote
19	80.5	44.0	448	E86291	hypothetical prote
20	80.5	44.0	829	S72366	DNA topoisomerase
21	80	43.7	144	T44863	trophozoite antige
22	80	43.7	356	T43145	hypothetical prote
23	80	43.7	373	AD0262	probable membrane
24	80	43.7	644	1KGHUH1	kininogen, HMW pre
25	79	43.2	65	D44863	trophozoite antige
26	79	43.2	353	AB2396	hypothetical prote
27	78	42.6	213	S04491	dermal gland prote
28	78	42.6	279	H90992	hypothetical prote
29	78	42.6	659	A36664	S59/2 homeotic pro

30	77.5	42.3	529	2	T50609	hypothetical prote
31	77.5	42.3	1560	2	T42727	proliferation pote
32	77	42.1	115	2	T51324	nickel-insertion a
33	77	42.1	231	2	AD0325	urease accessory p
34	77	42.1	270	2	A26480	knob protein-mal
35	77	42.1	473	2	A54494	knob-associated hi
36	77	42.1	634	2	A54495	knob protein precu
37	77	42.1	634	2	A28412	histidine-rich pro
38	77	42.1	654	2	B71623	knob-associated Hi
39	76	41.5	326	2	D83483	probable metal tra
40	76	41.5	400	2	S58222	PQ-rich protein -
41	76	41.5	1095	2	T00329	hypothetical prote
42	75.5	41.3	619	1	KGBOH2	kininogen, HMW II
43	75.5	41.3	621	1	KGBOH1	kininogen, HMW I p
44	75	41.0	196	2	G83138	hypothetical prote
45	75	41.0	250	2	B35026	filaggrin B - mous

ALIGNMENTS

RESULT 1

D64049

adhesin homolog HI0119 - Haemophilus influenzae (strain Rd KW20)

C;Species: Haemophilus influenzae

C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 21-Jul-2000

C;Accession: D64049

R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, .
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.L.; Glodek, A.; Kelley, J.M.; Weidman, .
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995

A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, .
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A;Reference number: A64000; MUID:95350630; PMID:7542800

A;Accession: D64049

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-337 <TIG>

A;Cross-references: GB:U32698; GB:L42023; NID:g3212178; PIDN:AAC21794.1; PID:g1573074;

C;Superfamily: hypothetical protein HI0119

Query Match 62.3%; Score 114; DB 2; Length 337;
Best Local Similarity 60.7%; Pred. No. 4.4e-06;
Matches 17; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 2 HKHKHKHKHKHKHKHKHKHKHKHKHKHK 29

Db 129 HKHKKHKHKHKHKHKHKHKHKHKHKHKHK 156

RESULT 2

S45444

BEH1 protein-binding protein BOB1 - yeast (Saccharomyces cerevisiae)

N;Alternate names: protein YBL0717; protein YBL085W

C;Species: Saccharomyces cerevisiae

C;Date: 09-Aug-1994 #sequence_revision 09-Sep-1994 #text_change 21-Jul-2000

C;Accession: S45444; S45421; S45826; S59218

R;Bender, A.; Bender, L.; Kokojan, V.

submitted to the EMBL Data Library, April 1994

A;Description: Yeast Boblp (Bemlp-binding protein) binds to the SH3 domain-containing p

A;Reference number: S45444

A;Accession: S45444

A;Molecule type: DNA

A;Residues: 1-980 <BEN>

A;Cross-references: EMBL:J31406; NID:g829041; PIDN:AA08439.1; PID:g466436

R;Obermaier, B.; Gassenhuber, J.; Piravandi, E.; Domdey, H.

submitted to the EMBL Data Library, May 1994

A;Description: Sequence analysis of a 78,6 kb segment of the left end of Saccharomyces c

A;Reference number: S45387

A;Accession: S45421

A;Molecule type: DNA

A;Residues: 1-980 <OBE>

A;Cross-references: EMBL:X79489; NID:g496661; PIDN:CAA56021.1; PID:g496694

Qy 2 HKHKHKHKHKHKHKHKHKHKHKHKHKHK 29

Db 124 HHHDDHEHHQDHDHDHDHDHDHEHHHHE 151

RESULT 9

128937
hypothetical protein C52B9.8 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T28937
R:Nelson, J.
submitted to the EMBL Data Library, July 1996
A:Description: The sequence of C. elegans cosmid C52B9.
A:Reference number: Z20545
A:Accession: T28937
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-1257 <NLE>
A:Cross-references: EMBL:U64598; PIDN:AAC47974.1; GSPDB:GN00028; CESP:C52B9.8
A:Experimental source: strain Bristol N2; clone C52B9
C:Genetics:
A:Gene: CESP:C52B9.8
A:Map position: X
A:Introns: 15/2; 321/3; 450/3; 596/2; 776/2; 823/2; 871/3

Query Match 45.9%; Score 84; DB 2; Length 1257;
Best Local Similarity 55.2%; Pred. No. 0.028;
Matches 16; Conservative 1; Mismatches 12; Indels 0; Gaps 0

Qy 1 KHKHKHKHKHKHKHKHKHKHKHKHKHKHK 29

Db 1152 KEPHKEKDKKEKEKEEDKGKEKHKDK 1180

RESULT 10

1SHUT1
DNA topoisomerase (EC 5.99.1.2) - human
N:Alternate names: nicking-closing enzyme; omega-protein; relaxing enzyme; swi
C:Species: Homo sapiens (man)
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 18-Jun-1999
C:Accession: A30887; A40008; S13821; S02397; S40643; A34422
R:D'Arpa, P.; Machlin, P.S.; Rattrie III, H.; Rothfield, N.F.; Cleveland, D.W.;
Proc. Natl. Acad. Sci. U.S.A. 85, 2543-2547, 1988
A:Title: cDNA cloning of human DNA topoisomerase I: catalytic activity of a 67
A:Reference number: A30887; MUID:98190108; PMID:2833744
A:Accession: A30887
A:Molecule type: mRNA
A:Residues: 1-765 <DAR>
A:Cross-references: GB:J03250; NID:G339805; PIDN:AAA61207.1; PID:G339806
R:Kunze, N.; Yang, G.; Doelberg, M.; Sundaip, R.; Knippers, R.; Richter, A.
J. Biol. Chem. 266, 9610-9616, 1991
A:Title: Structure of the human type I DNA topoisomerase gene.
A:Reference number: A40008; MUID:91236733; PMID:1851751
A:Accession: A40008
A:Molecule type: DNA
A:Residues: 1-144, 'A', 146-553, 'E', 555-765 <KUN>
A:Cross-references: GB:M60688; GB:M60689; GB:M60690; GB:M60691; GB:M60692; GB:M60693
M60704; GB:M60705; GB:M60706
R:Kunze, N.; Klein, M.; Richter, A.; Knippers, R.
Eur. J. Biochem. 194, 323-330, 1990
A:Title: Structural characterization of the human DNA topoisomerase I gene pro
A:Reference number: S13821; MUID:91099302; PMID:2176592
A:Accession: S13821
A:Molecule type: DNA
A:Residues: 1-20 <KU2>
A:Cross-references: EMBL:X52601
R:Oddou, P.; Schmidt, U.; Knippers, R.; Richter, A.
Eur. J. Biochem. 177, 523-529, 1988
A:Title: Monoclonal antibodies neutralizing mammalian DNA topoisomerase I activ
A:Reference number: S02397; MUID:89064806; PMID:2461859
A:Accession: S02397
A:Molecule type: mRNA

A;Residues: 344-765 <ODD>
A;Cross-references: GB:M60657
R;Tamura, H.O.; Kohchi, C.; Yamada, R.; Ikeda, T.; Koiwai, O.; Patterson, E.; Keene, J.D
Nucleic Acids Res. 19, 69-75, 1991
A;Title: Molecular cloning of a cDNA of a camptothecin-resistant human DNA topoisomerase
A;Reference number: S40643; MUID:91187651; PMID:1849260
A;Accession: S40643
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 523-543; 573-582, 'D', 584-593 <TAM>
R;Maul, G.G.; Jimenez, S.A.; Riggs, E.; Ziennicka-Kotula, D.
Proc. Natl. Acad. Sci. U.S.A. 86, 8492-8496, 1989
A;Title: Determination of an epitope of the diffuse systemic sclerosis marker antigen DN
ity in systemic sclerosis.
A;Reference number: A34422; MUID:90046823; PMID:2479024
A;Accession: A34422
A;Molecule type: mRNA
A;Residues: 657-765 <MAU>
A;Cross-references: GB:M27913; NID:G339807; PIDN:AAA61208.1; PID:G339808
C;Comment: Type I DNA topoisomerase catalyzes the ATP-independent transient breakage of
in another, followed by rejoining. This reaction will lead to the conversion of one top
C;Genetics:
A;Gene: GDB:TOP1
A;Cross-references: GDB:120444; OMIM:126420
A;Map position: 20q12-20q13.1
A;Introns: 11/3; 20/1
A;Note: the list of introns is incomplete
C;Superfamily: eukaryotic type I DNA topoisomerase
C;Keywords: DNA binding; DNA replication; heterotetramer; isomerase
F;723/Active site: Tyr #status predicted

Query Match 45.6%; Score 83.5; DB 1; Length 765;
Best Local Similarity 36.1%; Pred. No. 0.021; 6; Indels 33; Gaps 3;
Matches 22; Conservative 0; Mismatches

QY 2 HKHKHKH-----KGKH-----KHKHKHKGKH-----KHKHKH 28
DB 22 HKHKHKHDKRHHKHEKKEKREKSHNSEHDKSEKKEKTKHKDGSSEKHKDKH 81

QY 29 K 29
DB 82 K 82

RESULT 11
A54523
histidine-rich protein - Plasmodium lophurae (fragment)
C;Species: Plasmodium lophurae
C;Date: 15-Oct-1994 #sequence_revision 15-Oct-1994 #text_change 20-Aug-1999
C;Accession: A54523
R;Irving, D.O.; Cross, G.A.M.; Feder, R.; Wallach, M.
Mol. Biochem. Parasitol. 18, 223-234, 1986
A;Title: Structure and organization of the histidine-rich protein gene of Plasmodium lo
A;Reference number: A54523; MUID:86174893; PMID:3007981
A;Accession: A54523
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-140 <IRV>
A;Cross-references: GB:M15317; NID:G160331; PIDN:AAA29616.1; PID:G552196
C;Superfamily: plasmodium histidine-rich protein
C;Keywords: tandem repeat

Query Match 45.4%; Score 83; DB 2; Length 140;
Best Local Similarity 44.4%; Pred. No. 0.0057;
Matches 12; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 2 HKHKHKHKGKHKGKHKGKHKGKHKGKH 28
DB 9 HHHHHHHHHHHHHHHHHHHHHHHHHHH 35

RESULT 12
T13429

hypothetical protein 3088.6 - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: T13429
R;Murphy, L.; Harris, D.; Barrell, B.
submitted to the EMBL Data Library, April 1999
A;Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A;Reference number: 217668
A;Accession: T13429
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-744 <MUR>
A;Cross-references: EMBL:AL009195; NID:e1355203; PID:e1202214; PIDN:CAAL5704.1
C;Genetics:
A;Cross-references: FlyBase:FBgn0000377
A;Map position: X 178/3; 625/1
A;Introns: 49/3; 178/3; 625/1
A;Note: EG:3088.6

Query Match 45.1%; Score 82.5; DB 2; Length 744;
Best Local Similarity 44.0%; Pred. No. 0.027;
Matches 11; Conservative 8; Mismatches 3; Indels 3; Gaps 1;

QY 2 HKHKHKHKGKHKGKHKGKHKGKH 26
DB 130 HSHHQHQHQHQHQHQH---HQHRH 151

RESULT 13
JU0144
DNA topoisomerase (EC 5.99.1.2) - mouse
N;Alternate names: type I DNA topoisomerase
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 20-Jun-2000
C;Accession: JU0144
R;Koiwai, O.; Yasui, Y.; Sakai, Y.; Watanabe, T.; Ishii, K.; Yanagihara, S.; Andoh, T.
Gene 125, 211-216, 1993
A;Title: Cloning of the mouse cDNA encoding DNA topoisomerase I and chromosomal location
A;Reference number: JU0144; MUID:93216125; PMID:8096488
A;Accession: JU0144
A;Molecule type: mRNA
A;Residues: 1-767 <KOI>
A;Cross-references: GB:D10061; NID:G220617; PIDN:BAA00950.1; PID:G220618
C;Comment: This enzyme catalyzes the transient breakage and rejoining of single strands
C;Genetics:
A;Gene: TopI
A;Map position: 2-54.5
A;Superfamily: eukaryotic type I DNA topoisomerase
C;Keywords: DNA binding; isomerase
F;725/Active site: Tyr #status predicted

Query Match 45.1%; Score 82.5; DB 2; Length 767;
Best Local Similarity 34.9%; Pred. No. 0.027;
Matches 22; Conservative 0; Mismatches 6; Indels 35; Gaps 3;

QY 2 HKHKHKH-----KGKH-----KHKHKHKGKH-----KHKH 26
DB 22 HKHKHKHDKRHHKHEKKEKREKSHNSEHDKSEKKEKTKHKDGSSEKHKD 81

QY 27 KHK 29
DB 82 KHK 84

RESULT 14
A49546
DNA topoisomerase (EC 5.99.1.2) - Chinese hamster
C;Species: Cricetulus griseus (Chinese hamster)
C;Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 16-Jul-1999
C;Accession: A49546; S32697; S32698
R;Tanizawa, A.; Bertrand, R.; Kohlhagen, G.; Tabuchi, A.; Jenkins, J.; Pommier, Y.
J. Biol. Chem. 268, 25463-25468, 1993
A;Title: Cloning of Chinese hamster DNA topoisomerase I cDNA and identification of a sn

A;Reference number: A49546; MUID:94064611; PMID:8244980
A;Accession: A49546
A;Molecule type: mRNA
A;Residues: 1-767 <TA2>
A;Cross-references: EMBL:D21625; NID:G297078; PIDN:CAA79748.1; PID:G297079
A;Experimental source: DC3F cells
A;Note: This form is not camptothecin-resistant
R;Tanizawa, A.; Tabuchi, A.; Berstrand, R.; Pommier, Y.
submitted to the EMBL Data Library, February 1993
A;Reference number: S32697
A;Accession: S32697
A;Molecule type: mRNA
A;Residues: 1-504, 'S', 506-767 <TA2>
A;Cross-references: EMBL:D21624; NID:G297076; PIDN:CAA9747.1; PID:G297077
A;Experimental source: DC3F/C-10 cells
A;Note: This form is camptothecin-resistant
C;Superfamily: eukaryotic type I DNA topoisomerase
C;Keywords: isomerase

Query Match	45.1%	Score 82.5;	DB 2;	Length 767;
Best Local Similarity	34.9%	Pred. No. 0.027;		
Matches 22;	Conservative 0;	Mismatches 6;	Indels 35;	Gaps 3;

Qy	2	HKHKHKH-----KQKH-----KHKKHKKH-----KHKH 26		
Db	22	HKKDKHKHREHRRHKKDKKDKREKSKHNSHHKDKSEKHKHKEKENTKKHKGSSSEKHKD 81		
Qy	27	KHK 29		
Db	82	KHK 84		

RESULT 15

S66148

gene pipsqueak protein A short form - fruit fly (*Drosophila melanogaster*)

C:Species: *Drosophila melanogaster*

C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 21-Jul-2000

C:Accession: S66148

R:Weber, U.; Siegel, V.; Mlodzik, M.

EMBO J. 14, 6247-6257, 1995

A:Title: pipsqueak encodes a novel nuclear protein required downstream of seven-up for b

A:Reference number: S66148; MUID:96134923; PMID:8557044

A:Accession: S66148

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-535 <WEB>

A:Cross-references: EMBL:X90986; NID:gl149498; PIDN:CAA62473.1; PID:gl149499

C:Genetics:

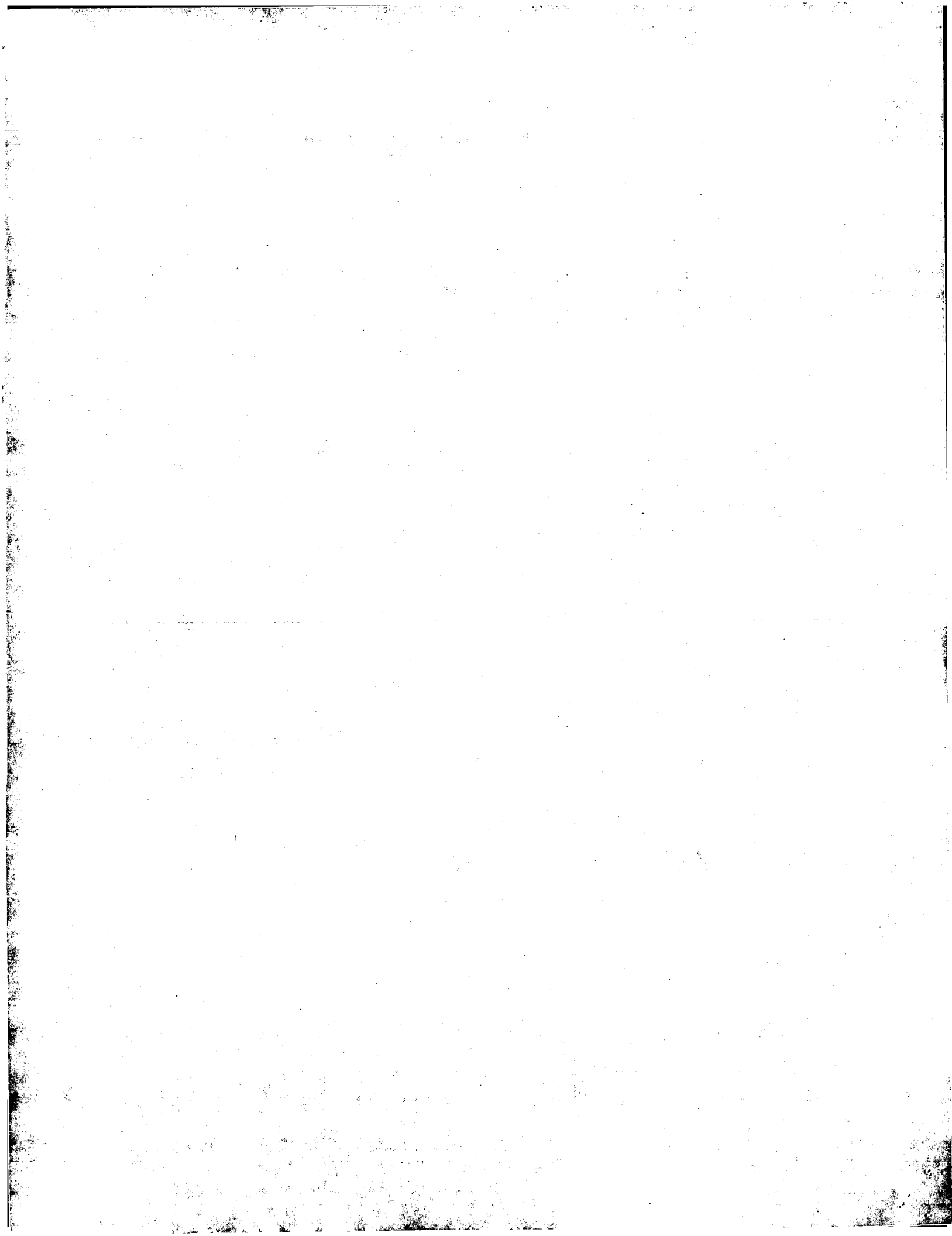
A:Gene: pipsqueak

C:Superfamily: POZ domain homology

F:21-123/domain: POZ domain homology <POZ>

```
Query Match      44.5%; Score 81.5; DB 2; Length 535;  
Best Local Similarity 39.4%; Pred. No. 0.036;  
Matches 13; Conservative 4; Mismatches 11; Indels 5; Gaps 1;
```

Search completed: January 20, 2004, 18:24:42
Job time : 15.7101 secs



RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schirml L.M., Staubli P., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Guetlich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Haegawa Y., Kawaji H., Kontseki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [3]
 RP SEQUENCE OF 185-1007 FROM N.A.
 RX MEDLINE=98369054; PubMed=9701556;
 RA Tate P., Lee M., Tweedie S., Skarnes W., Bickmore W.;
 RT "Capturing novel mouse genes encoding chromosomal and other nuclear
 RT proteins.";
 RL J. Cell Sci. 111:2575-2585(1998).
 RN [4]
 RP SEQUENCE OF 512-1007 FROM N.A.
 RX MEDLINE=97250302; PubMed=9102632;
 RA Gross T., Lutzberger M., Wiegmann H., Klingenhoff A., Shenoy S.,
 RA Kaeuter N.F.;
 RT "Functional analysis of the fission yeast Prp4 protein kinase
 RT involved in pre-mRNA splicing and isolation of a putative mammalian
 RT homologue.";
 RL Nucleic Acids Res. 25:1028-1035(1997).
 CC -1- FUNCTION: Has a role in pre-mRNA splicing. Phosphorylates SP2/ASF.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
 CC -1- SUBUNIT: Interacts with Clk1 C-terminus.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- PTM: Phosphorylated by Clk1.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF283466; AAM19102.1; -;
 CC EMBL; AK020579; BAB32137.1; -;
 CC EMBL; AK021274; BAB32358.1; -;
 CC EMBL; AF033663; AAC32042.1; -;
 CC EMBL; U48737; AAB03269.1; -;
 CC HSP; P24941; IAK1.
 CC MGD; MGI:109584; Prpf4b.
 CC GO; GO:0005694; C:Chromosome; IDA.
 CC InterPro; IPR000719; Prot kinase.
 CC InterPro; IPR002290; Ser Thr kinase.
 CC Pfam; PF00069; pkinase.1.
 CC ProDom; PD000001; Prot kinase; 1.
 CC SMART; SM00220; S_TKc; 1.
 CC SMART; SM00219; TyKc; 1.
 CC PROSITE; PS00107; PROTEIN KINASE ATP; FALSE_NEG.
 CC PROSITE; PS00108; PROTEIN_KINASE_ST.1.
 CC PROSITE; PS00011; PROTEIN_KINASE_DOM.1.
 KW mRNA processing; mRNA splicing; Transferase;
 KW Serine/threonine-protein kinase; ATP-binding; Phosphorylation;
 KW Nuclear protein.

FT DOMAIN 41 79 HIS-RICH.
 FT DOMAIN 40 497 ARG/LYS-RICH.
 FT DOMAIN 687 1006 PROTEIN KINASE.
 FT NP_BIND 693 701 ATP (BY SIMILARITY).
 FT BINDING 717 717 ATP (BY SIMILARITY).
 FT ACT_SITE 815 815 BY SIMILARITY.
 FT CONFLICT 185 187 SKS -> IFG (IN REF. 3).
 FT CONFLICT 223 223 K -> I (IN REF. 3).
 FT CONFLICT 633 633 F -> L (IN REF. 4).
 SQ SEQUENCE 1007 AA; 116947 MW; 18E6C3C43BE7AB4C CRC64;
 Query Match 60.4%; Score 110.5; DB 1; Length 1007;
 Best Local Similarity 59.0%; Pred. No. 5.6e-06;
 Matches 23; Conservative 1; Mismatches 4; Indels 11; Gaps 3;
 QY 1 KH-----KHKKHKKHKKHKKHKKG-----KHKKHKKHKK 29
 DB 40 KHSRHKKKKKHKKRSRHK-KHKKSSSEDRKKKKHKKHKK 77
 RESULT 4
 BOBL YEAST STANDARD; PRT; 980 AA.
 ID BOBL YEAST
 AC P38041;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE BOBL protein (BEM1-binding protein).
 GN BOBL OR BOIL OR YBL085W OR YBL0717.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96234675; PubMed=86666672;
 RA Bender L., Lo H.S., Lee H., Kokojan V., Peterson V., Bender A.;
 RT "Associations among PH and SH3 domain-containing proteins and
 RT Rho-type GTPases in Yeast.";
 RL J. Cell Biol. 133:879-894(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=S288c;
 RC MEDLINE=96076635; PubMed=7502586;
 RA Obermaier B., Gassenhuber J., Piravandi E., Domdey H.;
 RT "Sequence analysis of a 78.6 kb segment of the left end of
 RT Saccharomyces cerevisiae chromosome II.";
 RL Yeast 11:1103-1112(1995).
 CC -1- FUNCTION: BINDS TO THE BEM1 PROTEIN.
 CC -1- SIMILARITY: Contains 1 SH3 domain.
 CC -1- SIMILARITY: Contains 1 PH domain.
 CC -1- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L31406; AAB08439.1; -;
 CC EMBL; X79489; CAA56021.1; -;
 CC EMBL; Z35846; CAA84906.1; -;
 CC FIR; S45444; S45444.
 CC SGD; S0000181; BOIL.
 CC GO; GO:0005935; C:bud neck; IDA.
 CC GO; GO:0005543; P:phospholipid binding activity; IDA.
 CC GO; GO:000283; P:establishment of cell polarity (sensu Sacch. . .; IGI.
 CC GO; GO:0007266; P:Rho protein signal transduction; IGI.
 CC InterPro; IPR001849; PH.
 CC InterPro; IPR001660; SAM.
 CC InterPro; IPR001452; SH3.

DR Pfam; PF00169; PH; 1.
DR Pfam; PF00536; SAM; 1.
DR Pfam; PF00018; SH3; 1.
DR Pfam; PF00023; PH; 1.
DR SMART; SM00454; SAM; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00003; PH DOMAIN; 1.
DR PROSITE; PS0105; SAM DOMAIN; 1.
DR PROSITE; PS0002; SH3; 1.
KW SH3 domain. 13 77 SH3.
FT DOMAIN 228 292 SAM.
FT DOMAIN 776 895 PH.
FT DOMAIN 776 895 PH.
SQ SEQUENCE 980 AA; 109295 MW; 09F1DD1F9BF30F36 CRC64;
Query Match 52.5%; Score 96; DB 1; Length 980;
Best Local Similarity 83.3%; Pred. No. 0.00025;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
OY 9 KGKHKHKHKHKHKHKHKHK 26
DB 685 EGKHKHKHKHKHKHKHKHN 702
RESULT 5
HRPX PLALO STANDARD; PRT; 351 AA.
AC P04929;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histidine-rich glycoprotein precursor.
OS Plasmodium lophurae.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5853;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85061618; PubMed=6095114;
RA Ravetch J.V., Feder R., Pavlovic A., Blobel G.
RT "Primary structure and genomic organization of the histidine-rich protein of the malaria parasite Plasmodium lophurae."
RL Nature 312:615-620(1984).
CC -!- MISCELLANEOUS: IN THE INTRAERYTHROCYTIC STAGES OF DEVELOPMENT OF P. LOPHURAE IN DUCKS. THERE IS A SYNTHESIS OF A MAJOR PROTEIN THAT ACCUMULATES TO COMPRISE AT LEAST 50% OF THE CELLULAR MASS: THE HISTIDINE RICH PROTEIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; X01469; CAA25698.1; --
DR PIR; A22692; KGZQLH.
KW Malaria; Repeat; Glycoprotein; Signal.
FT SIGNAL 1 23
FT PROPEP 24 47 HISTIDINE-RICH GLYCOPROTEIN.
FT CHAIN 48 351 N-LINKED (GLCNAC. .) (PROBABLE).
FT CARBOHYD 40 40
FT DOMAIN 59 90 2 X 16 AA TANDEM REPEATS.
FT REPEAT 59 74 16-1.
FT REPEAT 75 90 16-2.
FT REPEAT 91 123 2 X 17 AA TANDEM REPEATS.
FT REPEAT 91 107 17-1.
FT REPEAT 108 123 17-2.
FT REPEAT 124 153 2 X 15 AA TANDEM REPEATS.
FT DOMAIN 124 138 15-1.
FT REPEAT 124 138 15-2.
FT REPEAT 139 153 18 X 10 AA TANDEM REPEATS.
FT DOMAIN 173 351 D19A48D47D890453 CRC64;
SQ SEQUENCE 351 AA; 44032 MW; 19A48D47D890453 CRC64;
Query Match 51.9%; Score 95; DB 1; Length 351;
Best Local Similarity 51.9%; Pred. No. 0.00012;
Matches 14; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
OY 2 HKHKHKHKHKHKHKHKHKHKHKHKHKHK 28
DB 215 HHHHHHHHHHHHHHHHHHHHHHHHHHHH 241
RESULT 6
BAF1 KLJMA STANDARD; PRT; 496 AA.
AC P33293;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-JUN-1994 (Rel. 23, Last annotation update)
DE Transcription factor BAF1 (ARS binding factor 1) (Protein ABF1) (Bidirectionally acting factor).
DE DE
GN ABF1.
OS Kluyveromyces marxianus (Yeast) (Kluyveromyces fragilis).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=4911;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CBS 6556;
RX MEDLINE=93277959; PubMed=7916634;
RA Oberly E.H.H., Maurer K., Mager W.H., Planta R.J.
RT "Structure of the ABF1-homologue from Kluyveromyces marxianus."
RL Biochim. Biophys. Acta 1173:233-236(1993).
CC -!- FUNCTION: TRANS-ACTING FACTOR IN THE REGULATION OF TRANSCRIPTION AND IN DNA REPLICATION. INVOLVED IN THE TRANSCRIPTION ACTIVATION OF A SUBSET OF RIBOSOMAL PROTEIN GENES. BINDS THE ARS-ELEMENTS FOUND IN MANY PROMOTERS. BINDS TO THE SEQUENCE 5'-TCN(7)ACG-3'.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- PTM: EXTENSIVELY PHOSPHORYLATED ON SER AND THR RESIDUES (BY SIMILARITY).
CC -!- SIMILARITY: STRONG TO YEAST BAF1, AND LOCAL TO YEAST RAP1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; Z19865; CAA79673.1; --
DR PIR; S33791; S33791.
DR InterPro; IPR006774; BAF1_ABFI.
DR Pfam; PF04684; BAF1_ABFI; 1.
DR Transcription regulation; Activator; DNA-binding; Nuclear protein;
KW Phosphorylation; Zinc-finger; Metal-binding; Zinc; DNA replication;
KW Trans-acting factor. 69
FT ZN FING 47 69 CHC2-TYPE.
FT DOMAIN 98 118 HIS-RICH.
FT DOMAIN 291 333 GLN/HIS-RICH.
FT DOMAIN 360 370 POLY-ALA.
FT MOD RES 402 402 PHOSPHORYLATION (BY PKC) (POTENTIAL).
FT SEQUENCE 496 AA; 55807 MW; 8F1E056D09C2E58F CRC64;
Query Match 50.8%; Score 93; DB 1; Length 496;
Best Local Similarity 34.5%; Pred. No. 0.00029;
Matches 10; Conservative 16; Mismatches 3; Indels 0; Gaps 0;
OY 1 KHKHKHKHKHKHKHKHKHKHKHKHKHKHK 29
DB 296 QHQHQHQHQHQHQHQHQHQHQHQHQHQHQ 324
RESULT 7
HYPB_BRAJA

```

OS Pasteurella haemolytica.
OC Bacterii; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Mannheimia.
OC NCBI_TaxID=75985;
RN [1];
RN SEQUENCE FROM N.A.
RP STRAIN=serotype A1 / ATCC 43270;
RP Graham M.R., Lo R.Y.C.;
RA Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RL -1- FUNCTION: INTERACTS WITH OUTER MEMBRANE RECEPTOR PROTEINS THAT
CC CARRY OUT HIGH-AFFINITY BINDING AND ENERGY DEPENDENT UPTAKE INTO
CC THE PERIPLASMIC SPACE OF SPECIFIC SUBSTRATES. IT COULD ACT TO
CC TRANSDUCE ENERGY FROM THE CYTOPLASMIC MEMBRANE TO SPECIFIC ENERGY-
CC REQUIRING PROCESSES IN THE OUTER MEMBRANE, RESULTING IN THE
CC RELEASE INTO THE PERIPLASM OF LIGANDS BOUND BY THESE OUTER
CC MEMBRANE PROTEINS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: ANCHORED TO THE CYTOPLASMIC
CC PERIPLASM VIA ITS N-TERMINAL SIGNAL-LIKE SEQUENCE, SPANS THE
CC PERIPLASM (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE TONB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/ or send an email to license@isb-sib.ch).
CC
CC EMBL; U62565; AAB09530.1; --
DR Transport; Protein transport; Inner membrane; Periplasmic;
KW Transmembrane; Signal-anchor.
FT DOMAIN 1 7 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 8 28 SIGNAL-ANCHOR (POTENTIAL).
FT DOMAIN 29 246 PERIPLASMIC (POTENTIAL).
FT SEQUENCE 246 AA; 27785 MW; C9582F619FCB5B5 CRC64;
SQ
Query Match 46.2%; Score 84.5; DB 1; Length 246;
Best Local Similarity 50.0%; Pred. No. 0.0014;
Matches 17; Conservative 2; Mismatches 8; Indels 7; Gaps 1;
Qy 2 HKGKHGKHK-----HGKHGKHGKHGKHGKH 28
Db 117 HKGHGHHKRELBQEQKPKPKPKRPHHHKH 150
RESULT 9
CC1L_HUMAN STANDARD; PRT; 726 AA.
ID CCT1_HUMAN
AC G060563; G060581;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cyclin T1 (Cyclin T) (CyclT).
GN CCNT1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1];
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=T-cell lymphoma;
RC MEDLINE=98150851; PubMed=9491887;
RA Wai P., Garber M.E., Fang S.-M., Fischer W.H., Jones K.A.;
RT "A novel CDK9-associated C-type cyclin interacts directly with HIV-1
RT Tat and mediates its high-affinity, loop-specific binding to TAR
RT RNA.";
RT Cell 92:451-462(1998).
RN [2];
RN SEQUENCE FROM N.A.
RP TISSUE=Breast cancer;
RA MEDLINE=98167917; PubMed=9499409;
RA Peng J.-M., Zhu Y., Milton J.T., Price D.H.;

```

PROSITE; PS00292; CYCLINS; FALSE NEG.
Cyclin; Cell cycle; Cell division; Coiled coil;
Transcription regulation; Nuclear protein.

FT DOMAIN 253..270 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
TAT-TAR RECOGNITION MOTIF (TRM).

FT FT 254..272 COILED COIL (POTENTIAL).
POLY-HIS.

FT FT 517..526 POLY-SER.
POLY-PRO.

FT FT 560..570 POLY-PRO.
C->Y; LOSS OF HIV-1 TAT TRANSACTIVATION.

FT FT 717..725 C->Y; R (IN REF. 2). Q -> R
MUTAGEN 261..281 Q -> R (IN REF. 2). CRC64;
CONFLICT 77..77 Q -> R (IN REF. 2). CRC64;
SQ SEQUENCE 726 AA; 80684 MW; 4637EFB2DDEDFE13 CRC64;

Query Match 45.9%; Score 84; DB 1; Length 726;
Best Local Similarity 51.7%; Pred No. 0.0045;
Matches 15; Conservative 1; Mismatches 9; Indels 4; Gaps 2;

DQ Y 1 KHKHKKK-KHGKHKKHHKKGKHKHKHX 28
::| | | | | | | | | | | | | | | |
Db 50S EHKKEKHTPSNHHHHNH--HSXKSH 530
|| || || || || || || || || || || || || || || ||

RESULT 10

CCCT1 HORSE

ID CCT1_HORSE STANDARD; PRT; 727 AA.

AC QSXTG;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
CYCLIN_T1 (Cylin T) (CyclT).

OS Equus caballus (Horse).
OC Chordata; Craniata; Vertebrata; Euteleostomi;
NCBI_taxid=9796;
[1]
SEQUENCE FROM N.A.
TISSUE=Fibroblast;
MEDLINE=99303751; PubMed=10373508;
Bioinformatics Institute.
Highly divergent lentiviral Tat proteins activate viral gene expression by a common mechanism." Mol. Cell. Biol. 19:4592-4598(1999)

-! FUNCTION: REGULATORY SUBUNIT OF THE CYCLIN-DEPENDENT KINASE PAIR (CDK9/CYLIN T) COMPLEX, ALSO CALLED POSITIVE TRANSCRIPTION ELONGATION FACTOR B (P-TEFB), WHICH IS PROPOSED TO FACILITATE THE TRANSITION FROM ABORTIVE TO PRODUCTION ELONGATION BY PHOSPHORYLATING THE CTD (CARBOXY-TERMINAL DOMAIN) OF THE LARGE SUBUNIT OF RNA POLYMERASE II (RNAP II) (BY SIMILARITY). BINDS ALSO TO THE TRANSCRIPTION DOMAIN OF THE EQUINE INFECTIOUS ANEMIA VIRUS (EIAV) NUCLEAR TRANSCRIPTIONAL ACTIVATOR, TAT. IT IS PROBABLY THE CIS-ACTING TRANSCRIPTION RESPONSE ELEMENT (TRE) RNA-BINDING COPROTOR FOR TAT. DOES NOT BIND TO THE TRANSCRIPTIONAL ACTIVITY OF THE HIV-1 NUCLEAR TRANSCRIPTIONAL ACTIVATOR.

-! SUBMIT: ASSOCIATES WITH CDK9 TO FORM P-TEFB. CYCLIN T1 IS THE PREDOMINANT CYCLIN ASSOCIATED WITH CDK9.

-! SUBCELLULAR LOCATION: Nuclear (By similarity).

-! SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN C SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement See http://www.isb-sib.ch/announce/ or send an email to licenses@isb-sib.ch

EMBL; AF137509; AAD38518.1; --
InterPro; IPR006670; Cyclin_N.
DR InterPro; IPR006671; Cyclin_N.
PFam; PF00134; cyclin_1.
SMART; SMART00385; CYCLIN; 1.
DR PROSITE; PS00292; CYCLINS; FALSE NEG.
DR

KW Cyclin; Cell cycle; Cell division; Coiled coil;
 KW Transcription regulation; Nuclear protein.
 FT DOMAIN 253 270 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 254 272 TAT-TAR RECOGNITION MOTIF (TRM) (BY
 FT SIMILARITY).
 FT DOMAIN 386 427 COILED COIL (POTENTIAL).
 FT DOMAIN 519 528 POLY-HIS.
 FT DOMAIN 562 573 POLY-SER.
 FT DOMAIN 718 726 POLY-PRO.
 SQ SEQUENCE 727 AA; 81013 MW; BPC2A398D6B35BCE CRC64;
 Query Match 45.9%; Score 84; DB 1; Length 727;
 Best Local Similarity 51.7%; Pred. No. 0.0045;
 Matches 15; Conservative 1; Mismatches 9; Indels 4; Gaps 2;
 QY 1 KHKKHK-KHGKHKHKHKHKHKHKH 28
 DB 507 EHKEKHKTHPSNHHHHHHH---HSHKHSH 532
 RESULT 11
 TOP1 HUMAN
 ID TOP1 HUMAN STANDARD; PRT; 765 AA.
 AC P11387; Q43256; Q12855; Q12856; Q9UUN0;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE DNA topoisomerase I (EC 5.99.1.2).
 GN TOP1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88190109; PubMed=2833744;
 RA D'Arpa P., Machlin P.S., Ratrie H. III, Rothfield N.F.,
 RA Cleveland D.W., Earnshaw W.C.;
 RT "CDNA cloning of human DNA topoisomerase I: catalytic activity of a
 RT 67.7-kDa carboxyl-terminal fragment";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:2543-2547(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91236733; PubMed=1851751;
 RA Kunze N., Yang G., Dolberg M., Sundarp R., Knippers R., Richter A.;
 RT "Structure of the human type I DNA topoisomerase gene";
 RL J. Biol. Chem. 266:9610-9616(1991).
 RN [3]
 RP SEQUENCE OF 5-765 FROM N.A., AND VARIANTS THR-370 AND SER-722.
 RC TISSUE=Peripheral blood;
 RX MEDLINE=95188170; PubMed=7882333;
 RA Fujimori A., Harker W.G., Kohlhaagen G., Hoki Y., Pommier Y.;
 RT "Mutation at the catalytic site of topoisomerase I in CEM/C2, a human
 RT leukemia cell line resistant to camptothecin";
 RL Cancer Res. 55:1339-1346(1995).
 RN [4]
 RP SEQUENCE OF 1-436 FROM N.A.
 RA Stuke C.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 541-765 FROM N.A.
 RX MEDLINE=89288043; PubMed=2544263;
 RA Zhou B.S., Bastow K.F., Cheng Y.C.;
 RT "Characterization of the 3' region of the human DNA topoisomerase I
 RT gene";
 RL Cancer Res. 49:3922-3927(1989).
 RN [6]
 RP SEQUENCE OF 657-765 FROM N.A.
 RX MEDLINE=90046823; PubMed=2479024;
 RA Maul G.G., Jimenez S.A., Riggs E., Ziennicka-Kotula D.;
 RT "Determination of an epitope of the diffuse systemic sclerosis marker
 RT antigen DNA topoisomerase I: sequence similarity with retroviral
 RT p30gag protein suggests a possible cause for autoimmunity in systemic

RT sclerosis.";

RL Proc. Natl. Acad. Sci. U.S.A. 86:8492-8496(1989).

RN [7]

RP VARIANTS CPT-RESISTANT.

RX MEDLINE=91187651; PubMed=1849260;

RA Tamura H., Kohchi C., Yamada R., Ikeda T., Koizumi O., Patterson E.,

RA Keene J.D., Okada K., Kjeldsen E., Nishikawa K.;

RT "Molecular cloning of a cDNA of a camptothecin-resistant human DNA

RT topoisomerase I and identification of mutation sites.";

RL Nucleic Acids Res. 19:69-75(1991).

RN [8]

RP VARIANT CPT-RESISTANT ALA-729.

RX MEDLINE=93075133; PubMed=1332703;

RA Kubota N., Kanawa F., Nishio K., Takeda Y., Ohmori T., Fujiwara Y.,

RA Terashima Y., Saijo N.;

RT "Detection of topoisomerase I gene point mutation in CPT-11 resistant

RT lung cancer cell line.";

RL Biochem. Biophys. Res. Commun. 188:571-577(1992).

RN [9]

RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 215-765.

RX MEDLINE=98155246; PubMed=9488644;

RA Redinbo M.R., Stewart L., Kuhn P., Champoux J.J., Hol W.G.J.;

RT "Crystal structures of human topoisomerase I in covalent and

RT noncovalent complexes with DNA.";

RL Science 279:1504-1513(1998).

RN [10]

RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 215-765.

RX MEDLINE=98155254; PubMed=9488652;

RA Stewart L., Redinbo M.R., Qiu X., Hol W.G.J., Champoux J.J.;

RT "A model for the mechanism of human topoisomerase I.";

RL Science 279:1534-1541(1998).

CC -I- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE

CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.

CC -I- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded

CC DNA, followed by passage and rejoining.

CC -I- ENZYME REGULATION: SPECIFICALLY INHIBITED BY CAMPTOTHECIN (CPT), A

CC PLANT ALKALOID WITH ANTITUMOR ACTIVITY.

CC -I- SUBUNIT: Monomer.

CC -I- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH

CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES

CC RELAX ONLY NEGATIVE SUPERCOILS.

CC -I- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA

CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN

CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS

CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND.

CC -I- SIMILARITY: BELONGS TO THE EUKARYOTIC TYPE I TOPOISOMERASE FAMILY.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; J03250; AAA61207.1; -

CC EMBL; M60706; AAA61206.1; -

CC EMBL; M60688; AAA61206.1; JOINED.

CC EMBL; M60689; AAA61206.1; JOINED.

CC EMBL; M60690; AAA61206.1; JOINED.

CC EMBL; M60691; AAA61206.1; JOINED.

CC EMBL; M60692; AAA61206.1; JOINED.

CC EMBL; M60693; AAA61206.1; JOINED.

CC EMBL; M60694; AAA61206.1; JOINED.

CC EMBL; M60695; AAA61206.1; JOINED.

CC EMBL; M60696; AAA61206.1; JOINED.

CC EMBL; M60697; AAA61206.1; JOINED.

CC EMBL; M60698; AAA61206.1; JOINED.

CC EMBL; M60699; AAA61206.1; JOINED.

CC EMBL; M60700; AAA61206.1; JOINED.

CC EMBL; M60701; AAA61206.1; JOINED.

CC EMBL; M60702; AAA61206.1; JOINED.

CC EMBL; M60703; AAA61206.1; JOINED.

```
EMBL; M60704; AAA61206.1; JOINED.
DR EMBL; M60705; AAA61206.1; JOINED.
DR EMBL; U07804; AAB60379.1; -.
DR EMBL; U07806; AAB60380.1; -.
DR EMBL; AL035852; CAB43980.1; -.
DR EMBL; X16479; CAA34500.1; ALT_INIT.
DR EMBL; M27913; AAA61208.1; -.
DR PIR; A30887; ISHUT1.
DR PDB; 1A31; 19-AUG-98.
DR PDB; 1A35; 26-AUG-98.
DR PDB; 1A36; 12-AUG-98.
DR PDB; 1EJ9; 03-AUG-00.
DR Aarhus/Ghent-2DPAGE; 610; NEPHGE.
DR Genew; HGNC:11986; TOP1.
DR GK; P11387; -.
DR MIM; 126420; -.
DR GO; GO:0003917; F:DNA topoisomerase I activity; TAS.
DR InterPro; IPR001631; Topoisomerase I.
DR Pfam; PF02919; Topoisomerase I_N; 1.
DR Pfam; PF01028; Topoisomerase I; 1.
DR PRINTS; PR00416; EUTPIISMRASEI.
DR SMART; SM00435; TOPEUC; 1.
DR PROSITE; PS00176; TOPOISOMERASE I EUK; 1.
DR isomerase; Topoisomerase; DNA-binding; Polymorphism; 3D-structure.
KW DOMAIN 191
FT ACT SITE 723 723
FT ACT SITE 370 370
FT VARIANT 533 533
FT VARIANT 722 722
FT VARIANT 729 729
FT CONFLICT 145 145
FT STRAND 220 222
FT STRAND 226 226
FT TURN 236 237
FT STRAND 240 242
FT TURN 243 244
FT STRAND 245 247
FT HELIX 251 261
FT TURN 262 265
FT HELIX 267 270
FT TURN 285 285
FT HELIX 272 284
FT TURN 285 285
FT HELIX 288 293
FT TURN 297 299
FT STRAND 300 301
FT HELIX 303 318
FT TURN 321 338
FT STRAND 340 343
FT TURN 344 345
FT STRAND 346 350
FT STRAND 354 354
FT STRAND 359 360
FT TURN 368 371
FT STRAND 373 374
FT HELIX 379 381
FT STRAND 383 385
FT TURN 388 389
FT TURN 397 398
FT STRAND 403 405
FT TURN 408 409
FT STRAND 414 417
FT TURN 419 421
FT STRAND 424 427
FT STRAND 429 429
FT TURN 431 432
FT HELIX 434 464
FT TURN 465 466
```

Query Match 45.6%; Score 83.5; DB 1; Length 765;
Best Local Similarity 36.1%; Pred. No. 0.0054; Indels 33; Gaps 3;
Matches 22; Conservative 0; Mismatches 6

QY 2 HKHKHKH-----KGKH-----KHKHKHKGH-----KHKHKH 28
DB 22 HKHKHKHDKRHRHKEHKEKDKSEKHSNSEKDKSEKHKKEKTKHKDGSSEKHKDKH 81
QY 29 K 29
DB 82 K 82

RESULT 12
TOP1 CRIGR STANDARD; PRT; 767 AA.
ID TOP1 CRIGR STANDARD; PRT; 767 AA.
AC Q07050;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA topoisomerase I (EC 5.99.1.2).
GN TOP1 OR TOP-1.
OS Cricetulus griseus (Chinese hamster).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OC Cricetulus.
OC NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94064611; PubMed=8244980;
RA Tanizawa A., Bertrand R., Kohlhaagen G., Tabuchi A., Jenkins J.,
Fommler Y.;
RT "Cloning of Chinese hamster DNA topoisomerase I cDNA and
RT identification of a single point mutation responsible for
RT camptothecin resistance."
RL J. Biol. Chem. 268:25463-25468 (1993).
CC -I- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE
CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.
CC -I- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded
CC DNA, followed by passage and rejoining
CC -I- ENZYME REGULATION: SPECIFICALLY INHIBITED BY CAMPTOTHECIN (CPT), A
CC PLANT ALKALOID WITH ANTITUMOR ACTIVITY.
CC -I- SUBUNIT: Monomer.
CC -I- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
CC RELAX ONLY NEGATIVE SUPERCOILS
CC -I- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA
CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK IN
CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS
CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND.
CC -I- SIMILARITY: BELONGS TO THE EUKARYOTIC TYPE I TOPOISOMERASE FAMILY.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; Z21624; CAA79747.1; -.
DR EMBL; Z21625; CAA79748.1; -.
DR PIR; A49546; A49546.
DR HSSP; P11387; 1A35.
DR InterPro; IPR001631; Topoisomerase I.
DR Pfam; PF02919; Topoisomerase I_N; 1.
DR Pfam; PF01028; Topoisomerase I; 1.
DR PRINTS; PR00416; EUTPIISMRASEI.
DR SMART; SM00435; TOPEUC; 1.

Pfam, PF01028; Topoisomerase I, 1.
 DR PRINTS; PR00416; EUTPISMRASE1.
 DR SMART; SM00435; TOPEUC; 1.
 DR PROSITE; PS00176; TOPOISOMERASE I EUK; 1.
 DR Isomerase; Topoisomerase; DNA-binding.
 FT ACT_SITE 725 725 DNA CLEAVAGE (BY SIMILARITY).
 FT CONFLICT 91 91 R -> P (IN REF. 2).
 FT CONFLICT 121 121 D -> E (IN REF. 2).
 FT CONFLICT 129 129 A -> V (IN REF. 2).
 FT CONFLICT 161 161 MISSING (IN REF. 2).
 FT CONFLICT 167 167 S -> L (IN REF. 2).
 FT CONFLICT 277 277 R -> W (IN REF. 2).
 FT CONFLICT 292 292 G -> E (IN REF. 2).
 FT CONFLICT 522 522 G -> V (IN REF. 2).
 FT CONFLICT 533 533 G -> W (IN REF. 2).
 FT CONFLICT 762 762 D -> Y (IN REF. 2).
 SQ SEQUENCE 767 AA; 590789 MW; 398327062B179F2A CRC64;

 Query Match 45.1%; Score 82.5; DB 1; Length 767;
 Best Local Similarity 34.9%; Pred.No. 0.007;
 Matches 22; Conservative 0; Mismatches 6; Indels 35; Gaps 3;

 QY 2 HKHKHKH-----KGKH-----KHKHKHKGKH-----KHKH 26
 Db 22 HKHKHKHDKREHHRHKEHKDKDKDKREKSKHSEHKDKSEKHKKEKTKHKGSSSEKHKD 81

 QY 27 KHK 29
 Db 82 KHK 84

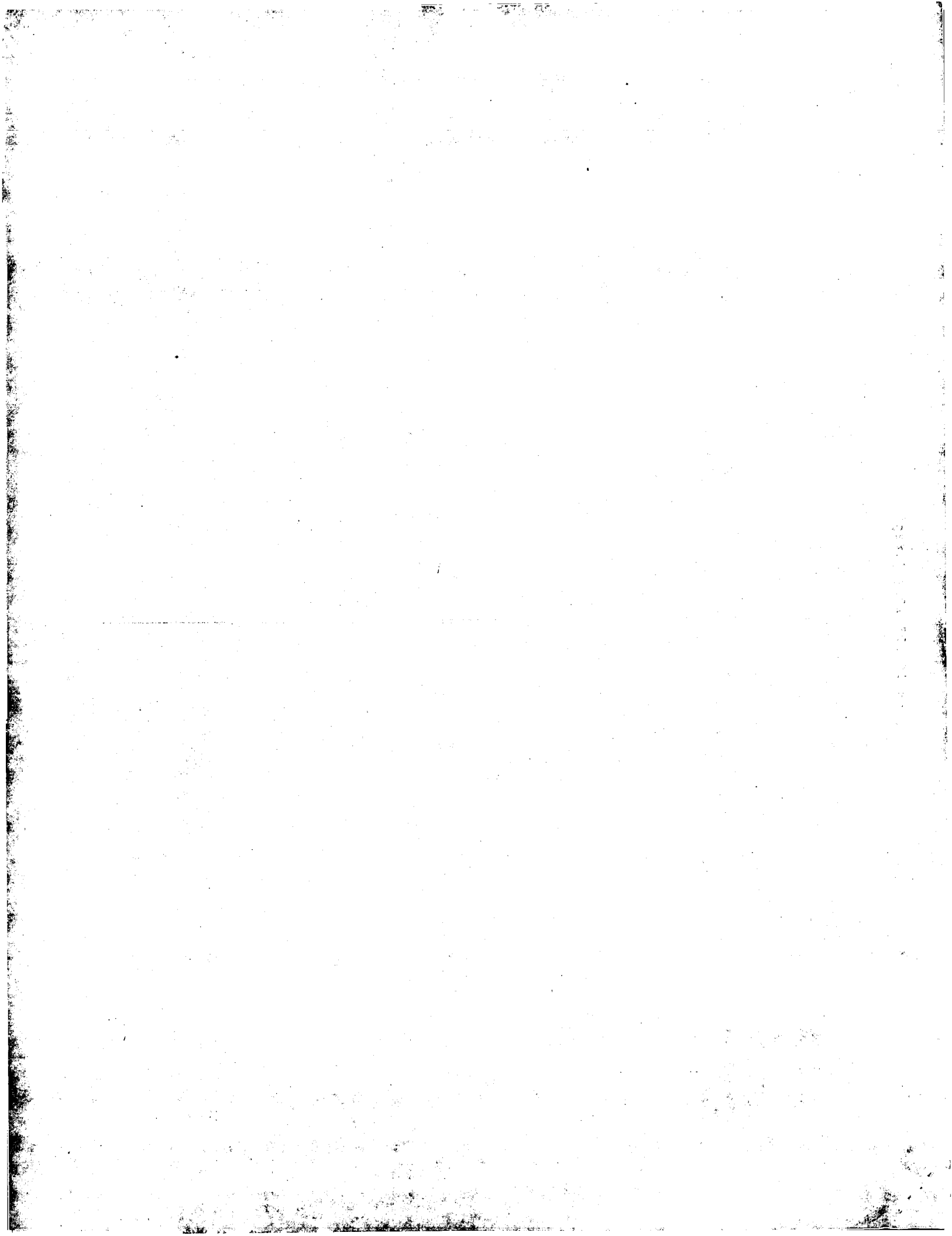
 RESULT 14
 T2D2 DROME STANDARD; PRT; 1221 AA.
 ID T2D2 DROME STANDARD; Q9VT64;
 AC Q24325; Q8SZR7; Q9VT64;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Transcription initiation factor TFIID 150 kDa subunit (TAFII-150)
 DE (TAFII150).
 OS TAF2 OR TAF150 OR CG6711.
 GN Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7227;
 RN [1]
 RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Embryo;
 RX MEDLINE=94233377; PubMed=8178153;
 RX Verrijzer C.P., Yokomori K., Chen J.-L., Tjian R.;
 RT "Drosophila TAFII150: similarity to yeast gene TSM-1 and specific
 RT binding to core promoter DNA";
 RL Science 264:933-941(1994).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Berkeley.
 RC MEDLINE=20196006; PubMed=10731132;
 RX Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RX Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G.G., Nelson C.R., Miklos G.L.G.,
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Prankoch C., Baldwin D.,
 RA Ballweg R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benton P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

[illegible]

SQ SEQUENCE 829 AA; 98230 MW; 8D1FE4252A916219 CRC64;
Query Match 44.0%; Score 80.5; DB 1; Length 829;
Best Local Similarity 41.9%; Pred. No. 0.013;
Matches 18; Conservative 2; Mismatches 8; Indels 15; Gaps 2;

Qy 2 HKHK-----HKHKGKHKKHKKHKKH-----KHKHKK 29
Db 22 HKHKDKKHKHKKHKKDKKDKREKSKHNNSEHRDPSEKHKDKHK 64

Search completed: January 20, 2004, 18:23:54
Job time : 10.6667 secs



Result No.	Score		Query		Length	DB	ID	Description
	Match	Match						
1	127	69.4	954	5	Q8ILV8	plasmodium		
2	119	65.0	3949	5	Q8IC24	plasmodium		
3	114	62.3	311	2	O68651	haemophilus		
4	113	61.7	366	16	Q9CMA8	pasteurella		
5	110.5	60.4	497	11	Q8C5G1	mus musculus		
6	110.5	60.4	991	4	Q9NQH2	homo sapien		
7	110.5	60.4	1007	4	Q8IVC3	homo sapien		
8	110.5	60.4	1007	11	Q8BND8	mus musculus		
9	108	59.0	776	5	O15793	plasmodium		
10	107	58.5	1461	5	Q8ST04	dictyosteli		
11	104.5	57.1	1189	4	O43604	homo sapien		
12	104.5	57.1	1199	4	O60668	homo sapien		
13	104.5	57.1	1199	4	O43487	homo sapien		
14	104	56.8	350	10	Q8GUC3	hordeum vul		
15	100	54.6	260	5	Q8SU46	encephalito		
16	99.5	54.4	83	17	Q8TIW6	methanosalic		

```

RESULT 4
QC9CMA8 PRELIMINARY; PRT; 366 AA.
ID Q9CWA8
Q9CWA8 AC Q9CWA8;
AC Q9CWA8; DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Fima.
GN FIMA OR PMO926.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteuriales;
OC Pastereuillaceae; Pasteurella.
NCBI_TaxId=747;
[1]
RN RN SEQUENCE FROM N.A.
RP STRAIN=Pm70;
RC MEDLINE=21145866; PubMed=11248100;
RX May B.J., Zhang O., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RA "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AAC006132; AAK03010.1; --
DR InterPro; IPR002395; Kininogen.
DR InterPro; IPR006128; Lipoprotein_4.
DR InterPro; IPR006127; SBP_bac_9.
DR Pfam; PF01297; SBP bac 9; 1..
DR PRINTS; PR00690; ADHESNFAMILY.
DR PRINTS; PR00334; KININOGEN.
DK Complete proteome.
KW COMPLETE PROTEOME.
SQ SEQUENCE 366 AA; 41443 MW; BF6A5929B78EC13D CRC64;

Query Match 61.7%; Score 113; DB 16; Length 366;
Best Local Similarity 63.0%; Pred. No. 3.Se-07;
Matches 17; Conservative 2; Mismatches 8; Indels 0; Gaps 0;


Qy 2 HKHKHKGHKGGKHKKHHKGKKHKKHX 28
Db 159 HXHDHDXDHXAHHKHGHEHKKHDEHKA 185


RESULT 5
QC5CGI PRELIMINARY; PRT; 497 AA.
ID QC5CGI
AC QC5CGI;
DC 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Pre-mRNA protein kinase (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mamalia; Euthera; Rodentia; Scurgnathi; Muridae; Murinae; Mus.
NCBI_TaxId=10090;
OX NCBI_TaxId=10090;
[1]
RN RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; Tissus=Sypathetic ganglion;
RC MEDLINE=22354683; PubMed=12466851;
RX The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK078611; BAC37340.1; -.
FT NON_TER 497
FT SEQUENCE 497 AA; 58250 MW; 3363B2738A0A1B66 CRC64;

Query Match 60.4%; Score 110.5; DB 11; Length 497;
Best Local Similarity 59.0%; Pred. No. 9.9e-07;
Matches 23; Conservative 4; Mismatches 4; Indels 11; Gaps 3


QY 1 KH-----KHKKHGKKGGKHKKHKG-----KHKKHKKKH 29
DB 40 KHSRHKCKKKHRSRK-KHSHSSBEDRGKKHKKHKKH 77

```

```
RESULT 6
Q9NQH2 PRELIMINARY; PRT; 991 AA.
AC Q9NQH2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE D1013A10.1 (PRP4 protein kinase homolog) (Fragment).
GN D1013A10.1
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kay M.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AL033383; CAB94780.1; -.
DR HSP; Q00534; IBI7.
DR InterPro; IPR000719; Prot kinase.
DR InterPro; IPR002290; Ser_Thr_kinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot kinase; 1.
DR SMART; SM00220; S_TKc; 1.
DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
FT NON TER 1
SQ SEQUENCE 991 AA; 115247 MW; F809E995B20671D2 CRC64;

Query Match 60.4%; Score 110.5; DB 4; Length 991;
Best Local Similarity 59.0%; Pred. No. 1.9e-06;
Matches 23; Conservative 1; Mismatches 4; Indels 11; Gaps 3;

QY 1 KH-----KHKKHKKGKHKKHKKG-----KHKKHKKHK 29
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 25 KHSRHKKKKKHRSKHK-KHKSSEEDKDKKKHKKHKKHKK 62

RESULT 7
Q8IVC3 PRELIMINARY; PRT; 1007 AA.
AC Q8IVC3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE PRP4 pre-mRNA processing factor 4 homolog B (Yeast).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC034969; AAH34969.1; -.
SQ SEQUENCE 1007 AA; 116990 MW; 5EEAF8AE8A7ACB9 CRC64;

Query Match 60.4%; Score 110.5; DB 4; Length 1007;
Best Local Similarity 59.0%; Pred. No. 2e-06;
Matches 23; Conservative 1; Mismatches 4; Indels 11; Gaps 3;

QY 1 KH-----KHKKHKKGKHKKHKKG-----KHKKHKKHK 29
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 39 KHSRHKKKKKHRSKHK-KHKSSEEDKDKKKHKKHKKHKK 76

RESULT 8
Q8BND8 PRELIMINARY; PRT; 1007 AA.
AC Q8BND8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Pre-mRNA protein kinase.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/60; TISSUE=Dorsal root ganglion;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK083926; BAC39069.1; -.
SQ SEQUENCE 1007 AA; 116975 MW; 18E1B7371E17AB4C CRC64;

Query Match 60.4%; Score 110.5; DB 11; Length 1007;
Best Local Similarity 59.0%; Pred. No. 2e-06;
Matches 23; Conservative 1; Mismatches 4; Indels 11; Gaps 3;

QY 1 KH-----KHKKHKKGKHKKHKKG-----KHKKHKKHK 29
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 40 KHSRHKKKKKHRSKHK-KHKSSEEDRDKKKHKKHKKHKK 77

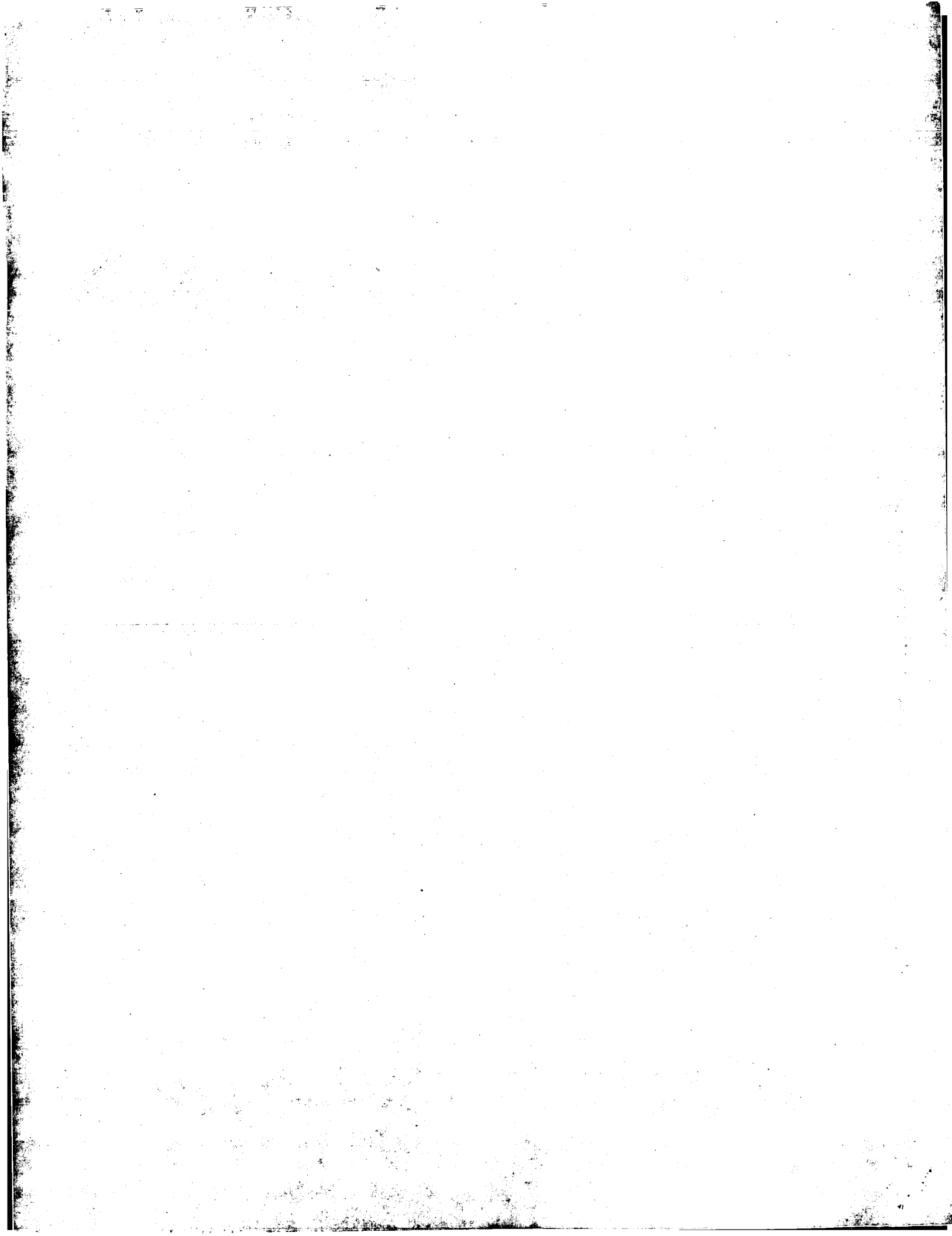
RESULT 9
Q15793 PRELIMINARY; PRT; 776 AA.
AC Q15793;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Strain Dd2 heat shock protein 86 (HSP86), O1 (o1), O3 (o3), O2 (o2),
DE CG8 (cg8), CG4 (cg4), CG3 (cg3), putative chloroquine resistance
DE transporter (crt), CG9 (cg9), CG1 (cg1), CG6 (cg6), CG2 (cg2), and CG7
DE (cg7) genes.
GN O1.
OS Plasmodium falciparum (isolate Dd2).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=57267;
RN [1]
RP SEQUENCE FROM N.A.
RA Su X., Kirkman L.A., Fujioka H., Wellem T.E.;
RX MEDLINE=98054002; PubMed=9393853;
RT "Complex polymorphisms in an approximately kDa protein are linked to
RT chloroquine-resistant P. falciparum in Southeast Asia and Africa.";
RL Cell 91:593-603(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=Dd2;
RX MEDLINE=21000495; PubMed=11090624;
RA Fidock D.A., Nomura T., Talley A.K., Cooper R.A., Dzekunov S.M.,
RA Ferdig M.T., Ursos L.M.B., Bir Singh Sidhu A., Naude B., Deitsch K.W.,
RA Su X.Z., Wootton J.C., Roepke P.D., Wellem T.E.;
RT "Mutations in the P. falciparum digestive vacuole transmembrane
RT protein PfCRT and evidence for their role in chloroquine resistance.";
RL Mol. Cell 6:861-871(2000).
DR EMBL; AF030694; AAC47848.1; -.
SQ SEQUENCE 776 AA; 92634 MW; 35F31748ABCA9EC CRC64;

Query Match 59.0%; Score 108; DB 5; Length 776;
Best Local Similarity 65.5%; Pred. No. 3.2e-06;
Matches 19; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 KHKKHKKGKHKKHKKHKKGKHKKHKKHKKHKKHKKHKK 29
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

2

Search completed: January 20, 2004, 18:26:21
Job time : 37.7246 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 18:05:34 ; Search time 31.0145 Seconds
(without alignments)
102.356 Million cell updates/sec

Title: US-10-018-103A-5

Perfect score: 136

Sequence: 1 KHKKHKKHKKHKKHKKHKK 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03:*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	136	100.0	20	AAE06230	Histidine copolymer
2	136	100.0	20	AAE06249	HH-K peptide seque
3	136	100.0	41	AAE06245	HH-K2b polymer. U
4	136	100.0	62	AAE06246	HH-K3b polymer. U
5	136	100.0	83	AAE06247	HH-K4b polymer. U
6	109	80.1	26	AAE06233	Histidine copolymer
7	105	77.2	15	AAE06240	Histidine copolymer
8	97	71.3	290	ABB68923	Drosophila melanog
9	94	69.1	227	ABG14399	Novel human diagno

10	94	69.1	1413	22	ABB60857	Drosophila melanog
11	94	69.1	1424	22	ABB60854	Drosophila melanog
12	93	68.4	49	22	ABG52113	Human liver peptid
13	93	68.4	49	22	ABB32046	Peptide #4697 enco
14	93	68.4	49	22	ABB37295	Peptide #4801 enco
15	93	68.4	49	22	ABB22584	Protein #4583 enco
16	93	68.4	49	22	AAAM57997	Human brain expres
17	93	68.4	49	22	AAAM70433	Human bone marrow
18	93	68.4	49	22	AAAM18263	Peptide #4697 enco
19	93	68.4	49	22	AAAM30755	Peptide #4792 enco
20	93	68.4	49	22	AAAM05877	Peptide #4559 enco
21	93	68.4	49	23	ABG40067	Human liver peptid
22	92	67.6	82	22	ABG47362	Human liver peptid
23	92	67.6	82	22	ABB27363	Human peptide #14
24	92	67.6	82	22	ABB32509	Peptide #15 encode
25	92	67.6	82	22	ABB18015	Protein #14 encode
26	92	67.6	82	22	AAAM53343	Human brain expres
27	92	67.6	82	22	AAAM65721	Human bone marrow
28	92	67.6	82	22	AAAM13580	Peptide #14 encode
29	92	67.6	82	22	AAAM25978	Peptide #15 encode
30	92	67.6	82	22	AAAM01333	Peptide #15 encode
31	92	67.6	82	23	ABG35350	Human peptide enco
32	92	67.6	96	20	AAW99830	HIV Vpr protein se
33	92	67.6	96	20	AAW99834	HIV A5P protein s
34	92	67.6	96	20	AAW99835	HIV L68S protein s
35	92	67.6	96	20	AAW99836	HIV H71C protein s
36	92	67.6	96	20	AAW99837	HIV G75A protein s
37	92	67.6	96	20	AAW99838	HIV C76S protein s
38	92	67.6	104	21	AAAB24812	Plant SDF encoded
39	92	67.6	104	21	AAAG44772	Arabidopsis thalia
40	92	67.6	134	22	ABG05192	Novel human diagno
41	92	67.6	183	22	ABG25337	Novel human diagno
42	92	67.6	195	21	AAAB24811	Plant SDF encoded
43	92	67.6	195	21	AAAG44771	Arabidopsis thalia
44	92	67.6	245	22	ABG28019	Novel human diagno
45	92	67.6	266	22	ABG03494	Novel human diagno

ALIGNMENTS

RESULT 1

AAE06230

ID AAE06230 standard; peptide; 20 AA.

XX

AC AAE06230;

XX

DT 25-SEP-2001 (first entry)

XX

DE Histidine copolymer, HH-K peptide.

XX

KW Histidine copolymer; enzyme deficiency; phenylalanine hydroxylase;
KW ornithine transcarbamylase; adenosine deaminase; tyrosine hydroxylase;
KW purine nucleoside phosphorylase; alpha1-antitrypsin; apolipoprotein E;
KW chronic granulomatous disease; sickle cell; beta-thalassemia; anaemia;
KW familial hypercholesterolaemia; low-density lipoprotein receptor;
KW phenylketonuria; Paconi's anaemia; haemophilia; muscular dystrophy;
KW cystic fibrosis; Parkinson's disease; retinitis pigmentosa; cytostatic;
KW lysosomal storage disease; mycoplasma pneumoniae type 1 disease; cardiant;
KW diabetic retinopathy; human immunodeficiency virus disease; infection;
KW cardiac disease; peripheral vascular disease; antisickling; arthritis;
KW nootropic; haemostatic; virucide; gene therapy.

XX Unidentified.

OS

XX WO200147496-A1.

PN

XX

PD 05-JUL-2001.

XX

PF 20-DEC-2000; 2000WO-US34603.

XX

PR 29-DEC-1999; 99US-0173576.

XX

PA (MIXS/) MIXSON A J.
 PI Mixson AJ;
 DR WPI; 2001-425579/45.
 XX Pharmaceutical composition useful for delivering therapeutic agent for
 PT treating diseases, comprises a peptide characterized with specified
 PT number of amino acids and a specified percentage of histidine residues
 PT -
 XX Claim 3; Page 36; 64pp; English.
 XX The invention relates to a pharmaceutical agent delivery composition
 XX comprising a transport polymer comprising a linear or branched peptide
 CC having about 10 to 300 amino acid residues, having about 5 to 100%
 CC histidine residues, and optionally having 1 to 95% non-histidine
 CC residues, at least one pharmaceutical agent and optionally one or more
 CC intracellular delivery components in association with the transport
 CC polymer. The pharmaceutical agent delivery composition is useful for
 CC delivering a pharmaceutical agent to the interior of a cell and is
 CC used in gene therapy for treating adenosine deaminase deficiency, purine
 CC nucleoside phosphorylase deficiency, chronic granulomatous disease with
 CC defective p47phox, sickle cell with HbS, beta-thalassemia due to
 CC inadequate production of beta-haemoglobin, familial hypercholesterolaemia
 CC due to a defective low-density lipoprotein receptor, phenylketonuria due
 CC to phenylalanine hydroxylase deficiency, alpha1-antitrypsin deficiency,
 CC ornithine transcarbamylase deficiency, Faconi's anaemia, apolipoprotein
 CC E deficiency, haemophilia A and B due to factor VIII and IX deficiency,
 CC muscular dystrophy due to dystrophin, laminin-2, or sacroglucans
 CC mutations, cystic fibrosis due to CFTR mutations, Parkinson due to
 CC tyrosine hydroxylase deficiency, retinitis pigmentosa, lysosomal storage
 CC disease (i.e., mycopolysaccharide type 1, Hunter, Hurler and Gaucher),
 CC diabetic retinopathy, human immunodeficiency virus disease, virus
 CC infection, acquired anaemia, cardiac and peripheral vascular disease
 CC and arthritis. The present peptide sequence is histidine copolymer
 CC HH-K.
 XX
 SQ Sequence 20 AA;
 Query Match 100.0%; Score 136; DB 22; Length 20;
 Best Local Similarity 100.0%; Pred. No. 1.1e-11;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KHKHKHKHKHKHKHKHKHK 20
 Db 1 KHKHKHKHKHKHKHKHKHK 20
 RESULT 2
 ID AAE06249 standard; peptide; 20 AA.
 XX
 AC AAE06249;
 XX
 DT 25-SEP-2001 (first entry)
 XX
 DE HH-K peptide sequence.
 XX
 KW Enzyme deficiency; phenylalanine hydroxylase;
 KW ornithine transcarbamylase; adenosine deaminase; tyrosine hydroxylase;
 KW purine nucleoside phosphorylase; alpha1-antitrypsin; apolipoprotein E;
 KW chronic granulomatous disease; sickle cell; beta-thalassemia; anaemia;
 KW familial hypercholesterolaemia; low-density lipoprotein receptor;
 KW phenylketonuria; Faconi's anaemia; haemophilia; muscular dystrophy;
 KW cystic fibrosis; Parkinson's disease; retinitis pigmentosa; cytostatic;
 KW lysosomal storage disease; mycopolysaccharide type 1 disease; cardiac;
 KW diabetic retinopathy; human immunodeficiency virus disease; infection;
 KW cardiac disease; peripheral vascular disease; antisickling; arthritis;
 KW neurotropic; haemostatic; virucide; gene therapy.
 XX Unidentified.
 OS
 XX

PN WO200147496-A1.
 XX
 PD 05-JUL-2001.
 XX
 PF 20-DEC-2000; 2000WO-US34603.
 XX
 PR 29-DEC-1999; 99US-0173576.
 XX
 PA (MIXS/) MIXSON A J.
 XX
 PI Mixson AJ;
 XX
 DR WPI; 2001-425579/45.
 XX
 PT Pharmaceutical composition useful for delivering therapeutic agent for
 PT treating diseases, comprises a peptide characterized with specified
 PT number of amino acids and a specified percentage of histidine residues
 PT -
 PS Disclosure; Fig 17; 64pp; English.
 XX The invention relates to a pharmaceutical agent delivery composition
 CC comprising a transport polymer comprising a linear or branched peptide
 CC having about 10 to 300 amino acid residues, having about 5 to 100%
 CC histidine residues, and optionally having 1 to 95% non-histidine
 CC residues, at least one pharmaceutical agent and optionally one or more
 CC intracellular delivery components in association with the transport
 CC polymer. The pharmaceutical agent delivery composition is useful for
 CC delivering a pharmaceutical agent to the interior of a cell and is
 CC used in gene therapy for treating adenosine deaminase deficiency, purine
 CC nucleoside phosphorylase deficiency, chronic granulomatous disease with
 CC defective p47phox, sickle cell with HbS, beta-thalassemia due to
 CC inadequate production of beta-haemoglobin, familial hypercholesterolaemia
 CC due to a defective low-density lipoprotein receptor, phenylketonuria due
 CC to phenylalanine hydroxylase deficiency, alpha1-antitrypsin deficiency,
 CC ornithine transcarbamylase deficiency, Faconi's anaemia, apolipoprotein
 CC E deficiency, haemophilia A and B due to factor VIII and IX deficiency,
 CC muscular dystrophy due to dystrophin, laminin-2, or sacroglucans
 CC mutations, cystic fibrosis due to CFTR mutations, Parkinson due to
 CC tyrosine hydroxylase deficiency, retinitis pigmentosa, lysosomal storage
 CC disease (i.e., mycopolysaccharide type 1, Hunter, Hurler and Gaucher),
 CC diabetic retinopathy, human immunodeficiency virus disease, virus
 CC infection, acquired anaemia, cardiac and peripheral vascular disease
 CC and arthritis. The present sequence is HH-K peptide sequence
 CC present in HH-k2b, HH-k3b and HHk4b polymers.
 XX
 SQ Sequence 20 AA;
 Query Match 100.0%; Score 136; DB 22; Length 20;
 Best Local Similarity 100.0%; Pred. No. 1.1e-11;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KHKHKHKHKHKHKHKHKHK 20
 Db 1 KHKHKHKHKHKHKHKHKHK 20
 RESULT 3
 ID AAE06245 standard; peptide; 41 AA.
 XX
 AC AAE06245;
 XX
 DT 25-SEP-2001 (first entry)
 XX
 DE HH-K2b polymer.
 XX Enzyme deficiency; phenylalanine hydroxylase;
 KW ornithine transcarbamylase; adenosine deaminase; tyrosine hydroxylase;
 KW purine nucleoside phosphorylase; alpha1-antitrypsin; apolipoprotein E;
 KW chronic granulomatous disease; sickle cell; beta-thalassemia; anaemia;
 KW familial hypercholesterolaemia; low-density lipoprotein receptor;
 KW phenylketonuria; Faconi's anaemia; haemophilia; muscular dystrophy;

KW cystic fibrosis; Parkinson's disease; retinitis pigmentosa; cytostatic;
 KW lysosomal storage disease; mycopolysaccharide type 1 disease; cardiac;
 KW diabetic retinopathy; human immunodeficiency virus disease; infection;
 KW cardiac disease; peripheral vascular disease; antiskinning; arthritis;
 KW neutropic; haemostatic; virucide; gene therapy.
 XX Unidentified.
 OS WO200147496-A1.
 PN 05-JUL-2001.
 PD 20-DEC-2000; 2000WO-US34603.
 PF 29-DEC-1999; 99US-0173576.
 XX (MIXS/) MIXSON A J.
 PA Mixson AJ;
 PI WPI; 2001-425579/45.
 DR Pharmaceutical composition useful for delivering therapeutic agent for
 XX treating diseases, comprises a peptide characterized with specified
 PT number of amino acids and a specified percentage of histidine residues
 PT -
 XX Example; Page 23; 64pp; English.
 PS The invention relates to a pharmaceutical agent delivery composition
 XX comprising a transport polymer comprising a linear or branched peptide
 CC having about 10 to 300 amino acid residues, having about 5 to 100%
 CC histidine residues, and optionally having 1 to 95% non-histidine
 CC residues, at least one pharmaceutical agent and optionally one or more
 CC intracellular delivery components in association with the transport
 CC polymer. The pharmaceutical agent delivery composition is useful for
 CC delivering a pharmaceutical agent to the interior of a cell and is
 CC used in gene therapy for treating adenosine deaminase deficiency, purine
 CC nucleoside phosphorylase deficiency, chronic granulomatous disease with
 CC defective p47phox, sickle cell with HbS, beta-thalassemia due to
 CC inadequate production of beta-haemoglobin, familial hypercholesterolaemia
 CC due to a defective low-density lipoprotein receptor, phenylketonuria due
 CC to phenylalanine hydroxylase deficiency, alpha1-antitrypsin deficiency,
 CC ornithine transcarbamylase deficiency, Faconi's anaemia, apolipoprotein
 CC E deficiency, haemophilia A and B due to factor VIII and IX deficiency,
 CC muscular dystrophy due to dystrophin, laminin-2, or sacroglucans
 CC mutations, cystic fibrosis due to CFTR mutations, Parkinson due to
 CC tyrosine hydroxylase deficiency, retinitis pigmentosa, lysosomal storage
 CC disease (i.e., mycopolysaccharide type 1, Hunter, Hurler and Gaucher),
 CC diabetic retinopathy, human immunodeficiency virus disease, virus
 CC infection, acquired anaemia, cardiac and peripheral vascular disease
 CC and arthritis. The present peptide sequence is HH-K2b polymer.
 XX Sequence 41 AA;
 SQ Query Match 100.0%; Score 136; DB 22; Length 41;
 Best Local Similarity 100.0%; Pred. No. 2.3e-11;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KHKKHKHKHKHKHKHKHKHK 20
 DB 1 KHKKHKHKHKHKHKHKHKHK 20
 RESULT 4
 AAE06246
 ID AAE06246 standard; peptide; 62 AA.
 XX AAE06246;
 AC
 XX 25-SEP-2001 (first entry)
 DT HH-K3b polymer.
 XX

XX Enzyme deficiency; phenylalanine hydroxylase;
 KW ornithine transcarbamylase; adenosine deaminase; tyrosine hydroxylase;
 KW purine nucleoside phosphorylase; alpha1-antitrypsin; apolipoprotein E;
 KW chronic granulomatous disease; sickle cell; beta-thalassemia; anaemia;
 KW familial hypercholesterolaemia; low-density lipoprotein receptor;
 KW phenylketonuria; Faconi's anaemia; haemophilia; muscular dystrophy;
 KW cystic fibrosis; Parkinson's disease; retinitis pigmentosa; cytostatic;
 KW lysosomal storage disease; mycopolysaccharide type 1 disease; cardiac;
 KW diabetic retinopathy; human immunodeficiency virus disease; infection;
 KW cardiac disease; peripheral vascular disease; antiskinning; arthritis;
 KW neutropic; haemostatic; virucide; gene therapy.
 XX Unidentified.
 OS WO200147496-A1.
 PN 05-JUL-2001.
 PD 20-DEC-2000; 2000WO-US34603.
 PF 29-DEC-1999; 99US-0173576.
 XX (MIXS/) MIXSON A J.
 PA Mixson AJ;
 PI WPI; 2001-425579/45.
 DR Pharmaceutical composition useful for delivering therapeutic agent for
 XX treating diseases, comprises a peptide characterized with specified
 PT number of amino acids and a specified percentage of histidine residues
 PT -
 XX Example; Page 23; 64pp; English.
 PS The invention relates to a pharmaceutical agent delivery composition
 XX comprising a transport polymer comprising a linear or branched peptide
 CC having about 10 to 300 amino acid residues, having about 5 to 100%
 CC histidine residues, and optionally having 1 to 95% non-histidine
 CC residues, at least one pharmaceutical agent and optionally one or more
 CC intracellular delivery components in association with the transport
 CC polymer. The pharmaceutical agent delivery composition is useful for
 CC delivering a pharmaceutical agent to the interior of a cell and is
 CC used in gene therapy for treating adenosine deaminase deficiency, purine
 CC nucleoside phosphorylase deficiency, chronic granulomatous disease with
 CC defective p47phox, sickle cell with HbS, beta-thalassemia due to
 CC inadequate production of beta-haemoglobin, familial hypercholesterolaemia
 CC due to a defective low-density lipoprotein receptor, phenylketonuria due
 CC to phenylalanine hydroxylase deficiency, alpha1-antitrypsin deficiency,
 CC ornithine transcarbamylase deficiency, Faconi's anaemia, apolipoprotein
 CC E deficiency, haemophilia A and B due to factor VIII and IX deficiency,
 CC muscular dystrophy due to dystrophin, laminin-2, or sacroglucans
 CC mutations, cystic fibrosis due to CFTR mutations, Parkinson due to
 CC tyrosine hydroxylase deficiency, retinitis pigmentosa, lysosomal storage
 CC disease (i.e., mycopolysaccharide type 1, Hunter, Hurler and Gaucher),
 CC diabetic retinopathy, human immunodeficiency virus disease, virus
 CC infection, acquired anaemia, cardiac and peripheral vascular disease
 CC and arthritis. The present peptide sequence is HH-K3b polymer.
 XX Sequence 62 AA;
 SQ Query Match 100.0%; Score 136; DB 22; Length 62;
 Best Local Similarity 100.0%; Pred. No. 3.6e-11;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KHKKHKHKHKHKHKHKHKHK 20
 DB 1 KHKKHKHKHKHKHKHKHKHK 20
 RESULT 5
 AAE06247

Query Match	100.0%;	Score 136;	DB 22;	Length 83;
Best Local Similarity	100.0%;	Pred. No. 4.8e-11;		
Matches 20;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

CC disease (i.e., mycopolysaccharide type 1, Hunter, Hurler and Gaucher),
 CC diabetic retinopathy, human immunodeficiency virus disease, virus
 CC infection, acquired anaemia, cardiac and peripheral vascular disease
 CC and arthritis. The present peptide sequence is histidine copolymer
 CC Y-HH.
 XX
 SQ Sequence 26 AA;
 Query Match 80.1%; Score 109; DB 22; Length 26;
 Best Local Similarity 83.3%; Pred. No. 4.5e-08;
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 KHKHKHKHKHKHKHKHK 18
 : : |||||
 Db 9 RRRHKHKHKHKHKHKHK 26
 RESULT 7
 AAE06240
 ID AAE06240 standard; peptide; 15 AA.
 AC AAE06240;
 XX
 XX 25-SEP-2001 (first entry)
 DT
 DE Histidine copolymer, peptide #3.
 XX
 XX Histidine copolymer; enzyme deficiency; phenylalanine hydroxylase;
 KW ornithine transcarbamylase; adenosine deaminase; tyrosine hydroxylase;
 KW purine nucleoside phosphorylase; algal-antitrypsin; apolipoprotein E;
 KW chronic granulomatous disease; sickle cell; beta-thalassaemia; anaemia;
 KW familial hypercholesterolaemia; low-density lipoprotein receptor;
 KW phenylketonuria; Faconi's anaemia; haemophilia; muscular dystrophy;
 KW cystic fibrosis; Parkinson's disease; retinitis pigmentosa; cytostatic;
 KW lysosomal storage disease; mycopolysaccharide type 1 disease; cardiant;
 KW diabetic retinopathy; human immunodeficiency virus disease; infection;
 KW cardiac disease; peripheral vascular disease; antisickling; arthritis;
 KW neutropic; haemostatic; virucide; gene therapy.
 XX
 OS Unidentified.
 XX
 XX WO200147496-A1.
 PN
 PD 05-JUL-2001.
 XX
 XX 20-DEC-2000; 2000WO-US34603.
 PF
 XX 29-DEC-1999; 99US-0173576.
 PR
 XX (MIXS/) MIXSON A J.
 PA
 XX Mixson AJ;
 PI
 XX WPI; 2001-425579/45.
 DR
 XX
 XX Pharmaceutical composition useful for delivering therapeutic agent for
 PT treating diseases, comprises a peptide characterized with specified
 PT number of amino acids and a specified percentage of histidine residues
 PT
 XX
 XX Claim 3; Page 36; 64pp; English.
 PS
 XX The invention relates to a pharmaceutical agent delivery composition
 CC comprising a transport polymer comprising a linear or branched peptide
 CC having about 10 to 300 amino acid residues, having about 5 to 100%
 CC histidine residues, and optionally having 1 to 95% non-histidine
 CC residues, at least one pharmaceutical agent and optionally one or more
 CC intracellular delivery components in association with the transport
 CC polymer. The pharmaceutical agent delivery composition is useful for
 CC delivering a pharmaceutical agent to the interior of a cell and is
 CC used in gene therapy for treating adenosine deaminase deficiency, purine
 CC nucleoside phosphorylase deficiency, chronic granulomatous disease with
 CC defective p47phox, sickle cell with HbS, beta-thalassaemia due to

CC inadequate production of beta-haemoglobin, familial hypercholesterolaemia
 CC due to a defective low-density lipoprotein receptor, phenylketonuria due
 CC to phenylalanine hydroxylase deficiency, algal-antitrypsin deficiency,
 CC ornithine transcarbamylase deficiency, Faconi's anaemia, apolipoprotein
 CC E deficiency, haemophilia A and B due to factor VIII and IX deficiency,
 CC muscular dystrophy due to dystrophin, laminin-2, or sacroglycans
 CC mutations, cystic fibrosis due to CFTR mutations, Parkinson due to
 CC tyrosine hydroxylase deficiency, retinitis pigmentosa, lysosomal storage
 CC disease (i.e., mycopolysaccharide type 1, Hunter, Hurler and Gaucher),
 CC diabetic retinopathy, human immunodeficiency virus disease, virus
 CC infection, acquired anaemia, cardiac and peripheral vascular disease
 CC and arthritis. The present peptide sequence is a histidine copolymer.
 XX
 SQ Sequence 15 AA;
 Query Match 77.2%; Score 105; DB 22; Length 15;
 Best Local Similarity 100.0%; Pred. No. 8.4e-08;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 HHKHKHKHKHKHKHKHK 18
 : : |||||
 Db 1 HHKHKHKHKHKHKHKHK 15
 RESULT 8
 ABB68923
 ID ABB68923 standard; Protein; 290 AA.
 XX
 AC ABB68923;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 33561.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 XX WO200171042-A2.
 PN
 PD 27-SEP-2001.
 XX
 XX 23-MAR-2001; 2001WO-US09231.
 PF
 XX 23-MAR-2000; 2000US-191637P.
 PR
 PR 11-JUL-2000; 2000US-0614150.
 XX
 XX (PEKE) PE CORP NY.
 PA
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI
 XX WPI; 2001-656860/75.
 DR
 DR N-PSDB; ABL13026.
 XX
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PT
 XX Disclosure; SEQ ID NO 33561; 21pp + Sequence Listing; English.
 PS
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (AB85737-AB872072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX

Wed Jan 21 11:27:50 2004

us-10-018-103a-5.rag

SQ Sequence 290 AA; Query Match 71.3%; Score 97; DB 22; Length 290; Best Local Similarity 68.4%; Pred. No. 1.8e-05; Mismatches 5; Indels 0; Gaps 0; Matches 13; Conservative 1; Mismatches 5; Indels 0; Gaps 0; Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 KKHKKHKKHKKHKKHKKH 19
 DB 118 KKHKKHKKHKKHKKHKKH 136

RESULT 9
 ID ABG14399 standard; Protein; 227 AA.
 AC ABG14399;
 DT 18-FEB-2002 (first entry)
 XX Novel human diagnostic protein #14390.
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 OS WO200175067-A2.
 PN 11-OCT-2001.
 PD 30-MAR-2001; 2001WO-US08631.
 PF 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 PA Drmanac RT, Liu C, Tang YT;
 PI WPI; 2001-639362/73.
 DR N-PSDB; RAS78586.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX Claim 20; SEQ ID No 44758; 103pp; English.
 PS The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 227 AA;

Query Match 69.1%; Score 94; DB 22; Length 227;
 Best Local Similarity 66.7%; Pred. No. 3.5e-05;
 Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 2 HKHKKHKKHKKHKKHKKH 19
 DB 141 HHHHHHHHHHHHHHHHHQH 158

RESULT 10
 ID ABB60857 standard; Protein; 1413 AA.
 AC ABB60857;
 DT 26-MAR-2002 (first entry)
 XX Drosophila melanogaster polypeptide SEQ ID NO 9363.
 DE Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX Drosophila melanogaster.
 OS WO200171042-A2.
 PN 27-SEP-2001.
 PD 23-MAR-2001; 2001WO-US09231.
 PF 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 DR N-PSDB; ABL04960.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX Disclosure; SEQ ID NO 9363; 21pp + Sequence Listing; English.
 PS The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABBS7737-ABBS73072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 1413 AA;

Query Match 69.1%; Score 94; DB 22; Length 1413;
 Best Local Similarity 63.2%; Pred. No. 0.00022;
 Matches 12; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 KKHKKHKKHKKHKKHKKH 19
 DB 1321 QHCHHMHSHHMHHAHPH 1339

RESULT 11
 ID ABB60854 standard; Protein; 1424 AA.
 XX

AC ABB60854;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 9354.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PP 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL04957.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 9354; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABBS7737-ABBS7072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1424 AA;
Query Match 69.1%; Score 94; DB 22; Length 1424;
Best Local Similarity 63.2%; Pred. No. 0.00023;
Matches 12; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 1 KHKKHKKHKKHKKHKKH 19
DB 1332 QHQHMHGSHHMHHAHPH 1350
RESULT 12
ABG52113
ID ABG52113 standard; Peptide; 49 AA.
AC ABG52113;
XX
DT 25-FEB-2003 (first entry)
XX
DE Human liver peptide, SEQ ID No 30761.
XX
KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
KW hypercholesterolaemia; coronary heart disease.
XX
OS Homo sapiens.
XX
PN WO200157273-A2.
XX

PD 09-AUG-2001.
XX
PP 30-JAN-2001; 2001WO-US00664.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488898/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analysing gene expression in human adult liver -
XX
PS Claim 27; SEQ ID No 30761; 658pp; English.
XX
CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult
CC liver. (I) may be used for predicting, measuring and displaying gene
CC expression in samples derived from human adult liver. The genes
CC identified may be involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which
CC is associated with coronary heart disease. ABG47348-ABG59930 represent
CC human liver single exon encoded peptides of the invention.
CC Note: the sequence information for this patent does not appear in the
CC printed specification but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 49 AA;
Query Match 68.4%; Score 93; DB 22; Length 49;
Best Local Similarity 66.7%; Pred. No. 9.8e-06;
Matches 12; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
QY 2 HKHKKHKKHKKHKKH 19
DB 3 HHHHHHHHHHHHHHHH 20
RESULT 13
ABB32046
ID ABB32046 standard; Peptide; 49 AA.
AC ABB32046;
XX
DT 01-FEB-2002 (first entry)
XX
DE Peptide #4697 encoded by breast cell single exon nucleic acid probe.
XX
KW Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer.
XX
OS Homo sapiens.
XX
PN WO200157271-A2.
XX
PD 09-AUG-2001.
XX
PP 30-JAN-2001; 2001WO-US00662.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
XX

Wed Jan 21 11:27:50 2004

PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-496933/54.
 XX New spatially-addressable set of single exon nucleic acid probes,
 XX useful for measuring gene expression in sample derived from human
 XX fetal liver, comprises number of single exon nucleic acid probes -
 XX Claim 27; SEQ ID NO 15014; 327pp + sequence listing; English.
 XX The invention relates to a spatially-addressable set of single exon
 XX nucleic acid probes for measuring gene expression in a sample derived
 XX from human breast and BT 474 cells. The method involves contacting
 XX the probes with a collection of detectably labelled nucleic acids
 XX derived from mRNA of human breast, and then measuring the label
 XX bound to each probe of the microarray. The probes are useful for
 XX verifying the expression of regions of genomic DNA predicted to
 XX encode proteins. They are useful for gene discovery, and for
 XX determining predisposition and/or prognosing breast disease. Gene
 XX expression analysis is useful for assessing the toxicity of chemical
 XX agents on cells. The microarray of this invention presents a far greater
 XX diversity of probes for measuring gene expression, with far less bias
 XX than expressed sequence tag microarrays. The method is suitable for
 XX rapid production of functional information from genomic sequence. The
 XX present sequence is a peptide encoded by a single exon nucleic acid
 XX probe of the invention.
 XX Note: The sequence data for this patent did not form part of the
 XX printed specification, but was obtained in electronic format directly
 XX from WPI at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 49 AA;
 Query Match 68.4%; Score 93; DB 22; Length 49;
 Best Local Similarity 66.7%; Pred. NO. 9.8e-06;
 Matches 12; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
 QY 2 HKHHKHHKHHKHHKHHK 19
 | | | | | | | | | | | | | | | | | | | | | |
 Db 3 HHHHHHHHHHHHHHHHHH 20
 RESULT 14
 ABB37295
 ID ABB37295 standard; Peptide; 49 AA.
 XX ABB37295;
 XX 04-FEB-2002 (first entry)
 XX Peptide #4801 encoded by human foetal liver single exon probe.
 XX Human; foetal liver; gene expression; single exon nucleic acid probe.
 XX Homo sapiens.
 XX WO200157277-A2.
 XX 09-AUG-2001.
 XX 30-JAN-2001; 2001WO-US00669.
 XX 04-FEB-2000; 2000US-0180312.
 XX 26-MAY-2000; 2000US-0207456.
 XX 30-JUN-2000; 2000US-0608408.
 XX 03-AUG-2000; 2000US-0632366.
 XX 21-SEP-2000; 2000US-0234687.
 XX 27-SEP-2000; 2000US-0236359.
 XX 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-496933/54.
 XX Single exon nucleic acid probes for analyzing gene expression in human
 XX hearts -

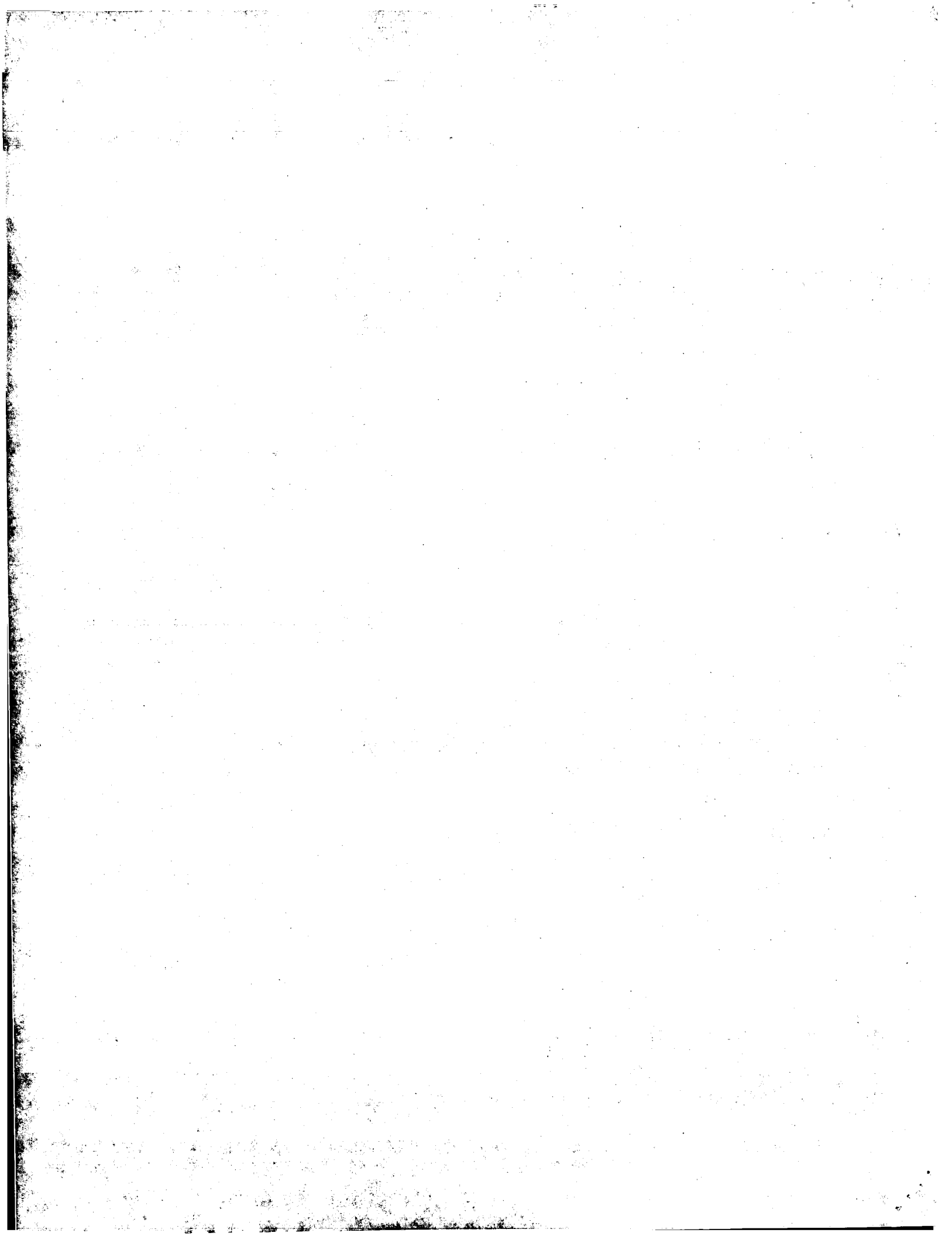
PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-483447/52.
 XX Human genome-derived single exon nucleic acid probes useful for
 XX analyzing gene expression in human fetal liver -
 XX Claim 27; SEQ ID NO 29930; 639pp + sequence listing; English.
 XX The invention relates to a single exon nucleic acid probe for
 XX measuring human gene expression in a sample derived from human foetal
 XX liver. The single exon nucleic acid probes may be used for predicting,
 XX measuring and displaying gene expression in samples derived from human
 XX fetal liver. The present sequence is a peptide encoded by a single exon
 XX nucleic acid probe of the invention.
 XX Note: The sequence data for this patent did not form part of the
 XX printed specification, but was obtained in electronic format directly
 XX from WPI at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 49 AA;
 Query Match 68.4%; Score 93; DB 22; Length 49;
 Best Local Similarity 66.7%; Pred. NO. 9.8e-06;
 Matches 12; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
 QY 2 HKHHKHHKHHKHHKHHK 19
 | | | | | | | | | | | | | | | | | | | | | |
 Db 3 HHHHHHHHHHHHHHHHHH 20
 RESULT 15
 ABB22584
 ID ABB22584 standard; Protein; 49 AA.
 XX ABB22584;
 XX 23-JAN-2002 (first entry)
 XX Protein #4593 encoded by probe for measuring heart cell gene expression.
 XX Human; gene expression; heart; microarray; vascular system;
 XX cardiovascular disease; hypertension; cardiac arrhythmia;
 XX congenital heart disease.
 XX Homo sapiens.
 XX WO200157274-A2.
 XX 09-AUG-2001.
 XX 30-JAN-2001; 2001WO-US00666.
 XX 04-FEB-2000; 2000US-0180312.
 XX 26-MAY-2000; 2000US-0207456.
 XX 30-JUN-2000; 2000US-0608408.
 XX 03-AUG-2000; 2000US-0632366.
 XX 21-SEP-2000; 2000US-0234687.
 XX 27-SEP-2000; 2000US-0236359.
 XX 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488899/53.
 XX Single exon nucleic acid probes for analyzing gene expression in human
 XX hearts -

XX Claim 15; SEQ ID No 24354; 530pp; English.
PS
XX The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart (see
CC ABA21535-ABA41305). The present sequence is a protein encoded by one such
CC probe. The probes may be used for predicting, measuring and displaying
CC gene expression in samples derived from the human heart via microarrays.
CC By measuring gene expression, the probes are useful for predicting,
CC diagnosing, grading, staging, monitoring and prognosing diseases of the
CC human heart and vascular system e.g. cardiovascular disease.
CC hypertension, cardiac arrhythmias and congenital heart disease.
CC Note: the sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 49 AA;
Query Match 68.4%; Score 93; DB 22; Length 49;
Best Local Similarity 66.7%; Pred. No. 9.8e-06;
Matches 12; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 HKHHKHKHKHKHKHKH 19
| | | | | | | | | |
Db 3 HHHHHHHHHHHHHHHH 20

Search completed: January 20, 2004, 18:23:18
Job time : 31.0145 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 18:22:04 ; Search time 11.3043 Seconds
(without alignments)
74.858 Million cell updates/sec

Title: US-10-018-103A-5
Perfect score: 136
Sequence: 1 KHKKHKKHKKHKKHKKHKK 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	92	67.6	363	4	US-09-328-352-4930
2	84	61.8	148	4	US-09-461-325-453
3	82.5	60.7	1199	3	US-09-208-742-2
4	82.5	60.7	1199	4	US-09-332-295-4
5	82.5	60.7	1199	4	US-09-709-979-4
6	79.5	58.5	763	2	US-08-677-862-2
7	79.5	58.5	763	2	US-09-252-571-2
8	79.5	58.5	763	3	US-09-434-065-2
9	79.5	58.5	763	3	US-08-789-275-4
10	79.5	58.5	763	3	US-08-789-275-5
11	78	57.4	303	2	US-08-203-532F-2
12	78	57.4	303	3	US-09-078-465-2
13	78	57.4	303	5	PCT-US95-01882A-2
14	77.5	57.0	60	1	US-08-255-437-1
15	77.5	57.0	60	1	US-09-115-032-1
16	77.5	57.0	60	5	PCT-US95-05772-1
17	76	55.9	115	4	US-09-461-325-160
18	74.5	54.8	434	4	US-09-252-991A-30855
19	74	54.4	115	4	US-09-996-243-95
20	74	54.4	633	3	US-08-557-006C-43
21	73.5	54.0	218	4	US-09-252-991A-25291
22	73.5	54.0	726	3	US-09-126-980-2
23	73.5	54.0	726	3	US-09-476-482-2
24	73.5	54.0	726	4	US-09-517-605-6
25	72.5	53.3	368	1	US-08-211-942-17
26	72	52.9	302	2	US-08-203-532F-4
27	72	52.9	302	3	US-08-950-860-16

28	72	52.9	302	3	US-09-078-465-4	Sequence 4, Appli
29	72	52.9	302	5	PCT-US95-01882A-4	Sequence 4, Appli
30	72	52.9	313	3	US-08-686-528A-3	Sequence 3, Appli
31	72	52.9	313	3	US-09-456-287-3	Sequence 3, Appli
32	72	52.9	337	3	US-08-686-528A-2	Sequence 2, Appli
33	72	52.9	337	3	US-09-456-287-2	Sequence 2, Appli
34	71	52.2	414	5	PCT-US92-06840-2	Sequence 2, Appli
35	71	52.2	1182	3	US-09-041-886-21	Sequence 21, Appli
36	71	52.2	2265	2	US-08-149-097D-36	Sequence 36, Appli
37	71	52.2	2509	2	US-08-149-097D-35	Sequence 35, Appli
38	70.5	51.8	38	1	US-08-287-717-8	Sequence 8, Appli
39	70.5	51.8	38	1	US-08-441-914-8	Sequence 8, Appli
40	69	50.7	480	1	US-07-882-292-2	Sequence 2, Appli
41	69	50.7	480	2	US-08-331-644-2	Sequence 2, Appli
42	69	50.7	480	5	PCT-US93-04102-2	Sequence 2, Appli
43	68	50.0	707	4	US-09-134-001C-2962	Sequence 2962, Ap
44	67.5	49.6	1257	1	US-08-340-4288-49	Sequence 49, Appli
45	67	49.3	189	1	US-08-152-922A-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-328-352-4930
; Sequence 4930, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4930
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4930

Query Match 67.6%; Score 92; DB 4; Length 363;
Best Local Similarity 66.7%; Pred. No. 7.2e-05;
Matches 12; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 HKHKKHKKHKKHKKHKKHKK 19
DB 188 HHHHHHHHHHHHHHHHH 205

RESULT 2
US-09-461-325-453
; Sequence 453, Application US/09461325A
; Patent No. 6475753
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 94 Human Secreted Proteins
; FILE REFERENCE: PZ029P1
; CURRENT APPLICATION NUMBER: US/09/461,325A
; CURRENT FILING DATE: 1999-12-14
; EARLIER APPLICATION NUMBER: PCT/US99/13418
; EARLIER FILING DATE: 1999-06-15
; EARLIER APPLICATION NUMBER: 60/089,507
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/089,508
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/089,509
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/089,510
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/090,112
; EARLIER FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: 60/090,113

Wed Jan 21 11:27:51 2004

```

; EARLIER FILING DATE: 1998-06-22
; NUMBER OF SEQ ID NOS: 532
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 453
; LENGTH: 148
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-461-325-453

Query Match      61.8%; Score 84; DB 4; Length 148;
Best Local Similarity 76.5%; Pred. No. 0.0003;
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy      3 KHKHKKHKKHKKHKKH 19
Db      119 KHKHKKHKKHKKHKKH 135

RESULT 3
US-09-208-742-2
; Sequence 2, Application US/09208742
; Patent No. 6174679
; GENERAL INFORMATION:
; APPLICANT: Kaufmann, Joerg
; TITLE OF INVENTION: C1F150/hTAF1150 is Necessary for Cell
; FILE REFERENCE: 1453.002
; CURRENT APPLICATION NUMBER: US/09/208,742
; CURRENT FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1199
; TYPE: PRT
; ORGANISM: human
; US-09-208-742-2

Query Match      60.7%; Score 82.5; DB 3; Length 1199;
Best Local Similarity 55.6%; Pred. No. 0.0033;
Matches 15; Conservative 1; Mismatches 4; Indels 7; Gaps 2;

Qy      1 KHKHKKHKKHKKH-----HKH-HKHK 20
Db      1142 KHSDDHHHHHHHEHKKKKKKHKKHKKH 1168

RESULT 4
US-09-332-295-4
; Sequence 4, Application US/09332295
; Patent No. 6303372
; GENERAL INFORMATION:
; APPLICANT: Kaufmann, Joerg
; TITLE OF INVENTION: C1F130 INHIBITS CELL CYCLE PROGRESSION
; FILE REFERENCE: 200130.456 / 1513.003
; CURRENT APPLICATION NUMBER: US/09/332,295
; CURRENT FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1199
; TYPE: PRT
; ORGANISM: Homo sapien
; US-09-332-295-4

Query Match      60.7%; Score 82.5; DB 4; Length 1199;
Best Local Similarity 55.6%; Pred. No. 0.0033;
Matches 15; Conservative 1; Mismatches 4; Indels 7; Gaps 2;

Qy      1 KHKHKKHKKHKKH-----HKH-HKHK 20
Db      1142 KHSDDHHHHHHHEHKKKKKKHKKHKKH 1168

; EARLIER FILING DATE: 1998-06-22
; NUMBER OF SEQ ID NOS: 532
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 453
; LENGTH: 148
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-461-325-453

Query Match      61.8%; Score 84; DB 4; Length 148;
Best Local Similarity 76.5%; Pred. No. 0.0003;
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy      3 KHKHKKHKKHKKHKKH 19
Db      119 KHKHKKHKKHKKHKKH 135

RESULT 5
US-09-709-979-4
; Sequence 4, Application US/09709979
; Patent No. 6423822
; GENERAL INFORMATION:
; APPLICANT: Kaufmann, Joerg
; TITLE OF INVENTION: C1F130 INHIBITS CELL CYCLE PROGRESSION
; FILE REFERENCE: 200130.456 / 1513.003
; CURRENT APPLICATION NUMBER: US/09/709,979
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 09/332,295
; PRIOR FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1199
; TYPE: PRT
; ORGANISM: Homo sapien
; US-09-709-979-4

Query Match      60.7%; Score 82.5; DB 4; Length 1199;
Best Local Similarity 55.6%; Pred. No. 0.0033;
Matches 15; Conservative 1; Mismatches 4; Indels 7; Gaps 2;

Qy      1 KHKHKKHKKHKKH-----HKH-HKHK 20
Db      1142 KHSDDHHHHHHHEHKKKKKKHKKHKKH 1168

RESULT 6
US-08-677-862-2
; Sequence 2, Application US/08677862
; Patent No. 5874230
; GENERAL INFORMATION:
; APPLICANT: SONG, Ho Yeong
; APPLICANT: ROTHE, Mike
; TITLE OF INVENTION: TRAF2-Associated Kinase
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, 34th floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/677,862
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brezner, David J
; REGISTRATION NUMBER: 24,774
; REFERENCE/DOCKET NUMBER: T96-005/A63613
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415 781-1989
; TELEFAX: 415 398-3249
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-677-862-2

Query Match      58.5%; Score 79.5; DB 2; Length 763;
Best Local Similarity 57.1%; Pred. No. 0.0051;
Matches 12; Conservative 0; Mismatches 6; Indels 3; Gaps 1;
```

Qy 2 HKHH---KHKHKHKHKHKH 19
| | | | | | | | | |
Db 599 HHHGNSHHHHHHHHHHH 619

RESULT 7
US-09-252-571-2
; Sequence 2, Application US/09252571
; Patent No. 5981250
; GENERAL INFORMATION:
; APPLICANT: SONG, Ho Yeong
; APPLICANT: ROTHE, Mike
; TITLE OF INVENTION: TRAF2-Associated Kinase
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, 34th floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/252,571
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/677,862
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brezner, David J
; REGISTRATION NUMBER: 24,774
; REFERENCE/DOCKET NUMBER: T96-005/A63613
; TELEPHONE: 415 781-1989
; TELEFAX: 415 398-3249
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; APPLICATION NUMBER: US/09/252,571

Query Match 58.5%; Score 79.5; DB 2; Length 763;
Best Local Similarity 57.1%; Pred. No. 0.0051;
Matches 12; Conservative 0; Mismatches 6; Indels 3; Gaps 1;
Qy 2 HKHH---KHKHKHKHKHKH 19
| | | | | | | | | |
Db 599 HHHGNSHHHHHHHHHHH 619

RESULT 8
US-09-434-065-2
; Sequence 2, Application US/09434065
; Patent No. 6107074
; GENERAL INFORMATION:
; APPLICANT: SONG, Ho Yeong
; APPLICANT: ROTHE, Mike
; TITLE OF INVENTION: TRAF2-Associated Kinase
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, 34th floor
; CITY: San Francisco
; STATE: CA

; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/434,065
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/677,862
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brezner, David J
; REGISTRATION NUMBER: 24,774
; REFERENCE/DOCKET NUMBER: T96-005/A63613
; TELEPHONE: 415 781-1989
; TELEFAX: 415 398-3249
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; APPLICATION NUMBER: US-09-434-065-2

Query Match 58.5%; Score 79.5; DB 3; Length 763;
Best Local Similarity 57.1%; Pred. No. 0.0051;
Matches 12; Conservative 0; Mismatches 6; Indels 3; Gaps 1;
Qy 2 HKHH---KHKHKHKHKHKH 19
| | | | | | | | | |
Db 599 HHHGNSHHHHHHHHHHH 619

RESULT 9
US-08-789-275-4
; Sequence 4, Application US/08789275A
; Patent No. 6251664
; GENERAL INFORMATION:
; APPLICANT: Palreja, Estivill Xavier
; APPLICANT: Pritchard, Melanie
; APPLICANT: Vilario, Jordi Guimera
; TITLE OF INVENTION: Human Gene Sequence of the Down Syndrome Critical
; Patent No. 6251664
; TITLE OF INVENTION: Region of Human Chromosome 21, Coding for A Serine-
; TITLE OF INVENTION: Threonine Protein Kinase (MNB), Expressed in the
; TITLE OF INVENTION: Neuronal Regions Affected in Down Syndrome
; FILE REFERENCE: U 01114-4
; CURRENT FILING DATE: 1997-01-28
; CURRENT APPLICATION NUMBER: US/08/789,275A
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 4
; LENGTH: 763
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
US-08-789-275-4

Query Match 58.5%; Score 79.5; DB 3; Length 763;
Best Local Similarity 57.1%; Pred. No. 0.0051;
Matches 12; Conservative 0; Mismatches 6; Indels 3; Gaps 1;
Qy 2 HKHH---KHKHKHKHKHKH 19
| | | | | | | | | |
Db 599 HHHGNSHHHHHHHHHHH 619

RESULT 10

us-10-018-103a-5.ra1

Wed Jan 21 11:27:51 2004

Query Match 57.4%; Score 78; DB 2; Length 303;
Best Local Similarity 55.6%; Pred. No. 0.0032;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 KHKHKKHKKHKKHKKHKK 20
DB 63 QHHRGHHHHHHHHHHHQ 80

RESULT 12
US-09-078-465-2
; Sequence 2, Application US/09078465
; Patent No. 6280969
; GENERAL INFORMATION:
; APPLICANT: Gorski, David H.
; APPLICANT: Walsh, Kenneth
; TITLE OF INVENTION: Growth Arrest Homeobox Gene
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calfee, Halter, and Griswold
; STREET: 800 Superior Avenue
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: U.S.A.
; ZIP: 44114-2688
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/078,465
; FILING DATE: 14-MAY-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/203,532
; FILING DATE: 24-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Golsick, Mary E.
; REGISTRATION NUMBER: 34829
; REFERENCE/DOCKET NUMBER: 22311/00114
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (216) 622-8200
; TELEFAX: (216) 241-0816
; TELEX: 980499
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; TYPE: amino acid
; LENGTH: 303 amino acids
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-078-465-2

Query Match 57.4%; Score 78; DB 3; Length 303;
Best Local Similarity 55.6%; Pred. No. 0.0032;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 KHKHKKHKKHKKHKKHKK 20
DB 63 QHHRGHHHHHHHHHHHQ 80

RESULT 13
PCT-US95-01882A-2
; Sequence 2, Application PC/TUS9501882A
; GENERAL INFORMATION:
; APPLICANT: Gorski, David H.
; APPLICANT: Walsh, Kenneth
; TITLE OF INVENTION: Growth Arrest Homeobox Gene
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calfee, Halter, and Griswold
; STREET: 800 Superior Avenue

US-08-789-275-5
; Sequence 5, Application US/08789275A
; Patent No. 6251664
; GENERAL INFORMATION:
; APPLICANT: Palleja, Estivill Xavier
; APPLICANT: Pritchard, Melanie
; APPLICANT: Vilario, Jordi Guimera
; TITLE OF INVENTION: Human Gene Sequence of the Down Syndrome Critical
; Patent No. 6251664
; TITLE OF INVENTION: Region of Human Chromosome 21, Coding for A Serine-
; TITLE OF INVENTION: Threonine Protein Kinase (MNB), Expressed in the
; TITLE OF INVENTION: Neuronal Regions Affected in Down Syndrome
; FILE REFERENCE: U 011114-4
; CURRENT APPLICATION NUMBER: US/08/789,275A
; CURRENT FILING DATE: 1997-01-28
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 5
; LENGTH: 763
; TYPE: PRT
; ORGANISM: Rat norvegicus
; FEATURE:
; US-08-789-275-5

Query Match 58.5%; Score 79.5; DB 3; Length 763;
Best Local Similarity 57.1%; Pred. No. 0.0051;
Matches 12; Conservative 0; Mismatches 6; Indels 3; Gaps 1;

QY 2 HKKH---KHKHKKHKKHKKH 19
DB 599 HHHGNSHHHHHHHHHHH 619

RESULT 11
US-08-203-532F-2
; Sequence 2, Application US/08203532F
; Patent No. 5856121
; GENERAL INFORMATION:
; APPLICANT: Gorski, David H.
; APPLICANT: Walsh, Kenneth
; TITLE OF INVENTION: Growth Arrest Homeobox Gene
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calfee, Halter, and Griswold
; STREET: 800 Superior Avenue
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: U.S.A.
; ZIP: 44114-2688
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/203,532F
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Golsick, Mary E.
; REGISTRATION NUMBER: 34829
; REFERENCE/DOCKET NUMBER: 22311/00114
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (216) 622-8200
; TELEFAX: (216) 241-0816
; TELEX: 980499
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 303 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-203-532F-2

;
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: U.S.A.
; ZIP: 44114-2688
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/01882A
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Golrick, Mary E.
; REGISTRATION NUMBER: 34829
; REFERENCE/DOCKET NUMBER: 22311/00114
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (216) 622-8200
; TELEFAX: (216) 241-0816
; TELEX: 980499
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 303 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-01882A-2

Query Match 57.4%; Score 78; DB 5; Length 303;
Best Local Similarity 55.6%; Pred. No. 0.0032;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 3 KHKHKKHKKHKKHKKHKK 20
Db 63 QHRRGHHHHHHHHHHQ 80

RESULT 14
US-08-255-457-1
; Sequence 1, Application US/08255457
; Patent No. 5780040
; GENERAL INFORMATION:
; APPLICANT: Plaut, Andrew G.
; APPLICANT: Gilbert-Rothstein, Joanne V.
; APPLICANT: Wright, Andrew
; TITLE OF INVENTION: HELICOBACTER PYLORI NICKEL BINDING
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/255,457
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/255,457
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul C.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00398/090001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 60 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-115-032-1

Query Match 57.0%; Score 77.5; DB 2; Length 60;
Best Local Similarity 52.2%; Pred. No. 0.0008;
Matches 12; Conservative 0; Mismatches 6; Indels 5; Gaps 1;

Qy 2 HKHKKHKKHKKHKKHKK 19
Db 11 HHHHHHHHHHHYHGGHHHHH 33

;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 60 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-255-457-1

Query Match 57.0%; Score 77.5; DB 1; Length 60;
Best Local Similarity 52.2%; Pred. No. 0.0008;
Matches 12; Conservative 0; Mismatches 6; Indels 5; Gaps 1;

Qy 2 HKHKKHKKHKKHKKHKK 19
Db 11 HHHHHHHHHHHYHGGHHHHH 33

RESULT 15
US-09-115-032-1
; Sequence 1, Application US/09115032
; Patent No. 5972348
; GENERAL INFORMATION:
; APPLICANT: Plaut, Andrew G.
; APPLICANT: Gilbert-Rothstein, Joanne V.
; APPLICANT: Wright, Andrew
; TITLE OF INVENTION: HELICOBACTER PYLORI NICKEL BINDING
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/115,032
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/255,457
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul C.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00398/090001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 60 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-115-032-1

Query Match 57.0%; Score 77.5; DB 2; Length 60;
Best Local Similarity 52.2%; Pred. No. 0.0008;
Matches 12; Conservative 0; Mismatches 6; Indels 5; Gaps 1;

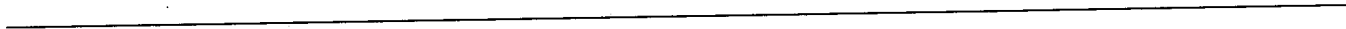
Qy 2 HKHKKHKKHKKHKKHKK 19
Db 11 HHHHHHHHHHHYHGGHHHHH 33

Search completed: January 20, 2004, 18:27:12

us-10-018-103a-5.ra1

Wed Jan 21 11:27:51 2004

Job time : 12.3043 secs



GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 18:23:25 ; Search time 22.6087 Seconds
(without alignments)
180.887 Million cell updates/sec

Title: US-10-018-103A-5

Perfect score: 136

Sequence: 1 KKHKKHHKHHKHHKHHK 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 762491 seqs, 204481190 residues

Total number of hits satisfying chosen parameters: 762491

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	136	100.0	20	12	US-10-131-909A-5
2	136	100.0	20	15	US-10-018-103A-5
3	109	80.1	26	12	US-10-131-909A-8
4	109	80.1	26	15	US-10-018-103A-8
5	105	77.2	15	12	US-10-131-909A-15
6	105	77.2	15	15	US-10-018-103A-15
7	94	69.1	378	12	US-10-029-386-33892
8	93	68.4	49	9	US-09-864-761-37882
9	92	67.6	82	9	US-09-864-761-33313
10	91	66.9	108	12	US-10-029-386-31185
11	90	66.2	78	9	US-09-864-761-37352
12	90	66.2	90	12	US-10-315-515-39
13	90	66.2	90	12	US-10-315-515-44
14	90	66.2	93	12	US-10-315-515-46
15	90	66.2	95	12	US-10-315-515-35

16	90	66.2	96	12	US-10-315-515-34	Sequence 34, Appl
17	90	66.2	96	12	US-10-315-515-36	Sequence 36, Appl
18	90	66.2	96	12	US-10-315-515-37	Sequence 37, Appl
19	90	66.2	96	12	US-10-315-515-40	Sequence 40, Appl
20	90	66.2	96	12	US-10-315-515-41	Sequence 41, Appl
21	90	66.2	96	12	US-10-315-515-42	Sequence 42, Appl
22	90	66.2	96	12	US-10-315-515-45	Sequence 45, Appl
23	90	66.2	105	12	US-10-315-515-43	Sequence 43, Appl
24	90	66.2	106	12	US-10-315-515-38	Sequence 38, Appl
25	89	65.4	87	9	US-09-864-761-33727	Sequence 33727, A
26	89	65.4	87	9	US-09-864-761-34744	Sequence 34744, A
27	89	65.4	461	10	US-09-764-868-765	Sequence 765, App
28	88	64.7	292	9	US-09-864-761-37944	Sequence 37944, A
29	86	63.2	16	12	US-10-104-307-17	Sequence 17, Appl
30	86	63.2	29	12	US-10-029-386-30014	Sequence 30014, A
31	86	63.2	1284	15	US-10-296-144-5	Sequence 5, Appli
32	84	61.8	148	15	US-10-012-542-453	Sequence 453, App
33	84	61.8	176	15	US-10-106-698-5906	Sequence 5906, Ap
34	84	61.8	186	12	US-10-029-386-34005	Sequence 34005, A
35	83	61.0	523	12	US-10-017-161-1982	Sequence 1982, Ap
36	83	61.0	523	12	US-10-292-798-1630	Sequence 1630, Ap
37	82.5	60.7	1199	14	US-10-147-268-4	Sequence 4, Appli
38	82.5	60.7	1199	15	US-10-338-279-4	Sequence 4, Appli
39	82	60.3	67	12	US-10-029-386-33229	Sequence 33229, A
40	82	60.3	156	9	US-09-864-761-41679	Sequence 41679, A
41	81.5	59.9	29	12	US-10-131-909A-4	Sequence 4, Appli
42	81.5	59.9	29	15	US-10-018-103A-4	Sequence 4, Appli
43	81.5	59.9	1007	15	US-10-211-133-7	Sequence 7, Appli
44	81.5	59.9	1043	15	US-10-097-340-258	Sequence 258, App
45	80	58.8	58	9	US-09-864-761-48985	Sequence 48985, A

ALIGNMENTS

RESULT 1

US-10-131-909A-5
; Sequence 5, Application US/10131909A
; Publication No. US20030165567A1
; GENERAL INFORMATION:
; APPLICANT: Mixson, A. James
; TITLE OF INVENTION: Histidine Copolymer and Methods For Using Same
; FILE REFERENCE: 5627*6
; CURRENT APPLICATION NUMBER: US/10/131,909A
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US 10/018103
; PRIOR FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: PCT/US00/34603
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/173576
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
US-10-131-909A-5

Query Match 100.0%; Score 136; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.5e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKHKKHHKHHKHHKHHKHK 20

Db 1 KKHKKHHKHHKHHKHHKHK 20

RESULT 2

US-10-018-103A-5
; Sequence 5, Application US/10018103A

Wed Jan 21 11:27:51 2004

```
; Publication No. US20030045465A1
; GENERAL INFORMATION:
; APPLICANT: Mixson, A. James
; TITLE OF INVENTION: Histidine Copolymer and Methods For Using Same
; FILE REFERENCE: 5627*5
; CURRENT APPLICATION NUMBER: US/10/018,103A
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US 60/173576
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: PCT/US00/34603
; PRIOR FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
US-10-018-103A-5
;
Query Match      80.1%; Score 136; DB 15; Length 26;
Best Local Similarity 100.0%; Pred. No. 6.5e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KHKHHKHHKHHKHHKHHK 20
Db 1 KHKHHKHHKHHKHHKHHK 20

RESULT 3
US-10-131-909A-8
; Sequence 8, Application US/10131909A
; Publication No. US20030165567A1
; GENERAL INFORMATION:
; APPLICANT: Mixson, A. James
; TITLE OF INVENTION: Histidine Copolymer and Methods For Using Same
; FILE REFERENCE: 5627*6
; CURRENT APPLICATION NUMBER: US/10/131,909A
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US 10/018103
; PRIOR FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: PCT/US00/34603
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/173576
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
US-10-131-909A-8
;
Query Match      80.1%; Score 109; DB 12; Length 26;
Best Local Similarity 100.0%; Pred. No. 6.4e-06;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KHKHHKHHKHHKHHKHHK 18
Db 9 RRRHHKHHKHHKHHKHHK 26

RESULT 4
US-10-018-103A-8
; Sequence 8, Application US/10018103A
; Publication No. US20030045465A1
; GENERAL INFORMATION:
; APPLICANT: Mixson, A. James
; TITLE OF INVENTION: Histidine Copolymer and Methods For Using Same
; FILE REFERENCE: 5627*5
; CURRENT APPLICATION NUMBER: US/10/018,103A
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US 60/173576
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
US-10-018-103A-8
;
Query Match      80.1%; Score 109; DB 12; Length 26;
Best Local Similarity 100.0%; Pred. No. 6.4e-06;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KHKHHKHHKHHKHHKHHK 18
Db 9 RRRHHKHHKHHKHHKHHK 26

RESULT 5
US-10-131-909A-15
; Sequence 15, Application US/10131909A
; Publication No. US20030165567A1
; GENERAL INFORMATION:
; APPLICANT: Mixson, A. James
; TITLE OF INVENTION: Histidine Copolymer and Methods For Using Same
; FILE REFERENCE: 5627*6
; CURRENT APPLICATION NUMBER: US/10/131,909A
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US 10/018103
; PRIOR FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: PCT/US00/34603
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/173576
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
US-10-131-909A-15
;
Query Match      77.2%; Score 105; DB 12; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 HHKHHKHHKHHKHHK 18
Db 1 HHKHHKHHKHHKHHK 15

RESULT 6
US-10-018-103A-15
; Sequence 15, Application US/10018103A
; Publication No. US20030045465A1
; GENERAL INFORMATION:
; APPLICANT: Mixson, A. James
; TITLE OF INVENTION: Histidine Copolymer and Methods For Using Same
; FILE REFERENCE: 5627*5
; CURRENT APPLICATION NUMBER: US/10/018,103A
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US 60/173576
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
US-10-018-103A-15
;
Query Match      77.2%; Score 105; DB 12; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 HHKHHKHHKHHKHHK 18
Db 1 HHKHHKHHKHHKHHK 15
```

; PRIOR FILING DATE: 2000-05-

; FILE REFERENCE: Aeomica-X-1

;
PRIOR FILING DATE: 2000-05-26

Query Match	66.9%;	Score 91;	DB 12;	Length 108;
Best Local Similarity	66.7%;	Pred. No. 0.0019;		
0. Mismatches				6: Indels

Qy

2	H K I N K H K H K H K H K H K H	19

p6

73	H P H H H H H H H H H H S H H H	90

RESULT 11

US-09-864-761-37352
; Sequence 37352, Application US/09864761-
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Derr Sharon G.

NUMBER OF SEQ ID NOS: 49117

```

1  SOFTWARE: Anomax Sequence Listing engine ver8. 1.1
2  SEQ ID NO 33313
3  LENGTH: 82
4  TYPE: PRT
5  ORGANISM: Homo sapiens
6  FEATURE:
7  OTHER INFORMATION: MAP TO AC007263.3
8  OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 8.3
9  OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7
10 OTHER INFORMATION: EXPRESSED IN FETAL, SIGNAL = 8.1
11 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.5
12 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.8
13 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 8.8
14 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 9.5
15 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.4
16 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.2
17 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 7.6
18 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.1
19 OTHER INFORMATION: SWISSPROT HIT: Q12329, EVALUAE 4.90e+00
20 US-09-864-761-33313

```

Query Match	67.6%	Score 92;	DB 9;	Length 82;
Best Local Similarity	66.7%	Pred. No. 0.0011;		
1. Mismatches				Indels
2. Gaps				
3. Mismatches				5;

Qy

2 HKHKHKHKHKHKHKH 19
| | | | | : |
p6

65 HHHHHHHHHHHYHRH 82

RESIT.T 10

RESULT 10
US-10-029-386-31185
; Sequence 31185, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:

```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 43117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37352
; LENGTH: 78
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF045448.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.5
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3
US-09-864-761-37352
```

```
Query Match 66.2%; Score 90; DB 9; Length 78;
Best Local Similarity 66.7%; Pred. No. 0.0018;
Matches 12; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 2 HKHKKHKKHKKHKKHKKH 19
   |||||
Db 44 HHHHHHHHHHHHHHHH 61
```

```
RESULT 12
US-10-315-515-39
; Sequence 39, Application US/10315515
; Publication No. US20030166190A1
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: NUCLEIC ACIDS RELATED TO PLANT
; TITLE OF INVENTION: RETROELEMENTS
; FILE REFERENCE: 08411-031001
; CURRENT APPLICATION NUMBER: US/10/315,515
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: US 60/339,060
; PRIOR FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 168
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Glycine max
US-10-315-515-39
```

```
Query Match 66.2%; Score 90; DB 12; Length 90;
Best Local Similarity 66.7%; Pred. No. 0.002;
Matches 12; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 2 HKHKKHKKHKKHKKHKKH 19
   |||||
Db 37 HHHHHHHHHHHHHHHH 54
```

```
RESULT 13
US-10-315-515-44
```

```
; Sequence 44, Application US/10315515
; Publication No. US20030166190A1
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: NUCLEIC ACIDS RELATED TO PLANT
; TITLE OF INVENTION: RETROELEMENTS
; FILE REFERENCE: 08411-031001
; CURRENT APPLICATION NUMBER: US/10/315,515
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: US 60/339,060
; PRIOR FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 168
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-315-515-44
```

```
Query Match 66.2%; Score 90; DB 12; Length 90;
Best Local Similarity 66.7%; Pred. No. 0.002;
Matches 12; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 2 HKHKKHKKHKKHKKHKKH 19
   |||||
Db 37 HHHHHHHHHHHHHHHH 54
```

```
RESULT 14
US-10-315-515-46
; Sequence 46, Application US/10315515
; Publication No. US20030166190A1
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: NUCLEIC ACIDS RELATED TO PLANT
; TITLE OF INVENTION: RETROELEMENTS
; FILE REFERENCE: 08411-031001
; CURRENT APPLICATION NUMBER: US/10/315,515
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: US 60/339,060
; PRIOR FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 168
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 46
; LENGTH: 93
; TYPE: PRT
; ORGANISM: Glycine max
US-10-315-515-46
```

```
Query Match 66.2%; Score 90; DB 12; Length 93;
Best Local Similarity 66.7%; Pred. No. 0.0021;
Matches 12; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 2 HKHKKHKKHKKHKKHKKH 19
   |||||
Db 43 HHHHHHHHHHHHHHHH 60
```

```
RESULT 15
US-10-315-515-35
; Sequence 35, Application US/10315515
; Publication No. US20030166190A1
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: NUCLEIC ACIDS RELATED TO PLANT
; TITLE OF INVENTION: RETROELEMENTS
; FILE REFERENCE: 08411-031001
; CURRENT APPLICATION NUMBER: US/10/315,515
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: US 60/339,060
```

us-10-018-103a-5.rapb

Wed Jan 21 11:27:51 2004

```
; PRIOR FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 168
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-315-515-35

Query Match      66.2%; Score 90; DB 12; Length 95;
Best Local Similarity 66.7%; Pred. No. 0.0021; 6; Indels 0; Gaps 0;
Matches 12; Conservative 0; Mismatches 6;

Cy 2 HHHHHHHHHHHHHHHHH 19
Db 43 HHHHHHHHHHHHHHHHH 60

Search completed: January 20, 2004, 18:28:43
Job time : 23.6087 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 18:18:04 ; Search time 10.1449 Seconds
(without alignments)
189.590 Million cell updates/sec

Title: US-10-018-103A-5

Perfect score: 136

Sequence: 1 KHKKHKHKHKHKHKHKHK

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_76:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	67.6	326	D83483	probable metal tra
2	91	66.9	1291	T13389	hypothetical prote
3	90	66.2	96	A44971	hypothetical prote
4	90	66.2	407	G84783	probable pectinest
5	89	65.4	139	B33910	sal homeotic prote
6	89	65.4	142	S00262	sal homeotic prote
7	89	65.4	142	C33910	sal homeotic prote
8	87	64.0	1040	T29092	TSC-22 protein hom
9	87	64.0	1212	T13804	shs protein - fru1
10	85	62.5	657	A29454	knob-associated hi
11	84	61.8	169	T34520	hypothetical prote
12	83	61.0	391	H86187	hypothetical prote
13	82.5	60.7	82	A29653	histidine-rich pro
14	82.5	60.7	385	A84696	probable zinc tran
15	82	60.3	351	KGZQHL	histidine-rich gly
16	81.5	59.9	387	A47446	HNF-3/fork head fa
17	81.5	59.9	451	A55909	transforming prote
18	81	59.6	658	T04219	hypothetical prote
19	79.5	58.5	385	OMRTSP	selenoprotein P pr
20	79.5	58.5	754	JC4898	Down-syndrome-crit
21	79	58.1	143	B64421	conserved hypothet
22	79	58.1	353	AB2396	hypothetical prote
23	78	57.4	303	A56837	homeobox protein M
24	78	57.4	303	B49122	growth arrest-spec
25	78	57.4	303	A48130	polycomb (PC) prot
26	78	57.4	390	A38565	neuron-derived rec
27	78	57.4	430	S66671	neuron-derived rec
28	78	57.4	625	S71930	neuron derived orp
29	78	57.4	628	JC2493	neuron derived orp

30 77.5 57.0 60 2 C64698 probable histidine
31 77.5 57.0 279 2 H90992 hypothetical prote
32 77 56.6 420 2 T39712 hypothetical prote
33 77 56.6 427 2 T42516 hypothetical prote
34 76 55.9 83 2 T16435 hypothetical prote
35 76 55.9 149 2 A54530 eggshell protein -
36 76 55.9 419 1 WJFPH2 homeotic protein H
37 76 55.9 529 2 T00677 hypothetical prote
38 76 55.9 580 2 T46024 hypothetical prote
39 75.5 55.5 283 2 C85838 hypothetical prote
40 75.5 55.5 474 2 A40721 neuroblast prolif
41 75.5 55.5 950 2 S27473 URBS1 protein - sm
42 75 55.1 274 2 A64978 hypothetical prote
43 75 55.1 351 2 T20270 hypothetical prote
44 75 55.1 434 2 I51436 HNF-3beta - Africa
45 75 55.1 529 2 T50609 hypothetical prote

ALIGNMENTS

RESULT 1

D83483

Probable metal transporter PA1297 [imported] - Pseudomonas aeruginosa (strain PAO1)
C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 06-Jan-2003

C:Accession: D83483

R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim
; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic path

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: D83483

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-326 <STO>

A:Cross-references: GB:AE004559; GB:AE004091; NID:g9947228; PIDN:AAG04686.1; GSPDB:GN00

A:Experimental source: strain PAO1

C:Genetics:

A:Gene: PA1297

C:Superfamily: zinc transporter ZnT-2

Query Match 67.6%; Score 92; DB 2; Length 326;
Best Local Similarity 63.2%; Pred. No. 0.00037;
Matches 12; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 2 KHKKHKHKHKHKHKHKHK 20
||| ||| ||| ||| |||
Db 163 HHHHHHHHHHHHHHHAHR 181

RESULT 2

T13389

hypothetical protein I15C2.10 - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000

C:Accession: T13389

R:Salles, C.; Valenti, P.; Darlamitsou, A.; Henderson, N.; Campbell, L.; Glover, D.
submitted to the EMBL Data Library, May 1999

A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.

A:Reference number: Z17665

A:Accession: T13389

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-1291 <CAT>

A:Cross-references: EMBL:AL031581; NID:e1320978; PID:e1426292; PIDN:CAA20894.2

C:Genetics:

A:Cross-references: FlyBase:FBgn0020381

A:Map position: X

A:Introns: 238/3; 1225/1

A>Note: EG:I15C2.10

A44971
Hypothetical protein 1 - Plasmodium brasilianum
C:Species: Plasmodium brasilianum
C:Date: 14-May-1993 #sequence_revision 14-May-1993 #text_change 17-Feb-1994
C:Accession: A44971
R:Kilejian, A.; Yang, Y.F.; Cochrane, A.H.; Rashid, M.A.
Mol. Biochem. Parasitol. 38, 291-293, 1990
A:title: Homologous sequences in Plasmodium cynomolgi and the gene of the histidine-rich
A:Reference number: A44971; MUID:90220761; PMID:2183051
A:Accession: A44971
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-96 <K11>
A:Cross-references: GB:M28064

RESULT 6
S00262
sal homeotic protein - fruit fly (*Drosophila melanogaster*)
Alternate names: homeotic protein spalt
C:Species: *Drosophila melanogaster*.
C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 20-Aug-1991
C:Accession: S00262; A38026; A33910
R:Frei, E.; Schuh, R.; Baumgartner, S.; Burri, M.; Noll, M.; Jürgens, G.
EMBO J. 7, 197-204, 1988
A:Title: Molecular characterization of spalt, a homeotic gene required for
A:Reference number: S00262

A:Residues: 1-142 <FR1>
A:Cross-references: EMBL:X57474; NID:G8536; PIDD:CAA40712.1; PID:G8537
A:Accession: A38026
A:Molecule type: mRNA
A:Residues: 1-142 <FR2>
R:Reuter, D.; Schuh, R.; Jaeckle, H.
P:Cell Natl. Acad. Sci. U.S.A. 86, 5483-5486, 1989
A>Title: The homeotic gene spalt (sal) evolved during Drosophila speciation
A:Reference number: A33910; MUID:89315821; PMID:2568636
A:Contents: annotation
C:Genetics:
A:Gene: sal
A:Cross-references: FlyBase:FBgn0003313
A:Map position: 2L 33A1.2
A:Introns: 4/3
C:Superfamily: homeotic protein sal
F:19-80/Region: glycine-rich
F:92-99,103-110/Region: 8-residue repeats
F:108-142/Region: histidine-rich

Query Watch	65.4%;	Score 89;	DB 2;	Length 142;
Best Local Similarity	64.7%;	Pred. No.	0.00038;	
Matches 11;	Conservative 4;	Mismatches 2;	Indels 0;	Gaps 0;
Qy	3 KHHKHHKKHKKHKKH	19		
	: : : :			
nb	126 EHHKHHGHEHGHHRH	142		

RESULT 7
C33910
sal homeotic protein - fruit fly (*Drosophila orena*)
C:Species: *Drosophila orena*
C>Date: 09-Mar-1990 #sequence_revision 09-Mar-1990 #text_change 21-Feb-1997
C:Accession: C33910
R:Reuter, D.; Schuh, R.; Jaeckle, H.
Proc. Natl. Acad. Sci. U.S.A. 86, 5483-5486, 1989
A:Title: The homeotic gene spalt (sal) evolved during *Drosophila* speciation.
A:Reference number: A33910; MUID:89315821; PMID:2568636
A:Accession: C33910

A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-142 <REU>
A;Cross-references: GB:M21579
C;Genetics:
A;Gene: FlyBase:Dore/sala
A;Cross-references: FlyBase:FBgn0012624
C;Superfamily: homeotic protein sal

Query Match 65.4%; Score 89; DB 2; Length 142;
Best Local Similarity 64.7%; Pred. No. 0.0038;
Matches 11; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 3 KHKHKHKHKHKHKHKH 19
|||:|||||:|||||:
Db 126 EHHEHGHGHEHGHHRH 142

RESULT 8
T29092
TSC-22 protein homolog - chicken
C;Species: Gallus gallus (chicken)
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T29092
R;Sawada, K.; Agata, K.; Eguchi, G.
submitted to the EMBL Data Library, December 1995
A;Description: Analysis of the cDNA library of chicken lens fibers..
A;Reference number: Z20561
A;Accession: T29092
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1040 <SAW>
A;Cross-references: EMBL:D82364; NID:d1067801; PID:d1012232; PIDN:BAAL1565.1
A;Experimental source: lens fibers

Query Match 64.0%; Score 87; DB 2; Length 1040;
Best Local Similarity 66.7%; Pred. No. 0.0039;
Matches 12; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 HKHKHKHKHKHKHKHKH 19
|||:|||||:|||||:
Db 200 HPHHVHHHHLHHHHHGH 217

RESULT 9
T13804
shs protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: T13804
R;Treisman, J.E.; Lai, Z.C.; Rubin, G.M.
Development 121, 2835-2845, 1995
A;Title: Shortighted acts in the decapentaplegic pathway in Drosophila eye development
A;Reference number: Z17767; MUID:96038094; PMID:7555710
A;Accession: T13804
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1212 <TRE>
A;Cross-references: EMBL:L42512; NID:g833709; PID:g833710; PIDN:AAC41608.1
C;Genetics:
A;Gene: shs
A;Cross-references: FlyBase:FBgn0010460

Query Match 64.0%; Score 87; DB 2; Length 1212;
Best Local Similarity 64.7%; Pred. No. 0.0045;
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 4 HKHKHKHKHKHKHKHKH 20
|||:|||||:|||||:
Db 306 HHHHHHHHHHHHHHHHGH 322

RESULT 10

A29454
knob-associated histidine-rich protein precursor - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 09-Jun-2000
C;Accession: A29454
R;Triglia, T.; Stahl, H.D.; Crewther, P.E.; Scanlon, D.; Brown, G.V.; Anders, R.F.; Kemp
EMBO J. 6, 1413-1419, 1987
A;Title: The complete sequence of the gene for the knob-associated histidine-rich protein
A;Reference number: A29454; MUID:87275836; PMID:3301325
A;Accession: A29454
A;Molecule type: DNA
A;Residues: 1-657 <TRI>
A;Cross-references: GB:Y00060; NID:g9908; PIDN:CAA68268.1; PID:g9909
C;Superfamily: knob-associated histidine-rich protein

Query Match 62.5%; Score 85; DB 2; Length 657;
Best Local Similarity 61.1%; Pred. No. 0.0044;
Matches 11; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 KHKHKHKHKHKHKHKHKH 20
|||:|||||:|||||:
Db 59 KQHCHHHHHHHHHHHQH 76

RESULT 11
T34520
hypothetical protein DKFZp564J157.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34520
R;Bloeker, H.; Boecher, M.; Brandt, P.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, October 1999
A;Reference number: Z21539
A;Accession: T34520
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-169 <BLO>
A;Cross-references: EMBL:AL122042
A;Experimental source: fetal brain; clone DKFZp564J157
C;Genetics:
A;Note: DKFZp564J157.1

Query Match 61.8%; Score 84; DB 2; Length 169;
Best Local Similarity 76.5%; Pred. No. 0.0017;
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 KHKHKHKHKHKHKHKHKH 19
|||:|||||:|||||:
Db 140 KMHKHKHKHKYHKGKH 156

RESULT 12
H86187
hypothetical protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 21-Jun-2002
C;Accession: H86187
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Lin, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luroo, J.S.; Maiti, R.; Marziani
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: H86187
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-391 <STO>
A;Cross-references: GB:AE005172; NID:g2388565; PIDN:AAB71446.1; GSPDB:GN00141

Wed Jan 21 11:27:52 2004

us-10-018-103a-5.rpr

R:Ravetch, J.V.; Feder, R.; Pavlovic, A.; Blobel, G.
Nature 312, 616-620, 1984
A:Title: Primary structure and genomic organization of the histidine-rich protein of the
A:Reference number: A22692; MUID:85061618; PMID:6095114
A:Superfamily: pectinesterase pemb
Query Match 61.0%; Score 83; DB 2; Length 391;
Best Local Similarity 68.8%; Pred. No. 0.0047; 4; Indels 0; Gaps 0;
Matches 11; Conservative 1; Mismatches 0; Gaps 0;
QY 5 HKHKKHKKHKKHKK 20
DB 47 HHHHHHHHHHHKK 62
RESULT 13
A29553
histidine-rich protein - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 09-Jun-2000
C:Accession: A29653
R:Lenstra, R.; d'Auriol, L.; Andrieu, B.; Le Bras, J.; Galibert, F.
Biochem. Biophys. Res. Commun. 146, 368-377, 1987
A:Title: Cloning and sequencing of Plasmodium falciparum DNA fragments containing repeti
A:Reference number: A90134; MUID:87270765; PMID:3038111
A:Accession: A29653
A:Molecule type: DNA
A:Residues: 1-82 <LEN>
A:Cross-references: GB:M17028; NID:g160339; PID:g160341
Query Match 60.7%; Score 82.5; DB 2; Length 82;
Best Local Similarity 57.9%; Pred. No. 0.0013;
Matches 11; Conservative 4; Mismatches 3; Indels 1; Gaps 1;
QY 2 HKHKKHKKHKKHKKHKK 20
DB 22 HHHH-HHHHHHHHHHQ 39
RESULT 14
A84696
probable zinc transporter [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: A84696
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: A84696
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-385 <STO>
A:Cross-references: GB:AE002093; NID:g3980394; PIDN:AAC95197.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g29410
A:Map position: 2
Query Match 60.7%; Score 82.5; DB 2; Length 385;
Best Local Similarity 59.1%; Pred. No. 0.0052;
Matches 13; Conservative 1; Mismatches 5; Indels 3; Gaps 1;
QY 2 HKHKKHKKHKKHKKHKK 20
DB 196 HHHHHHHHHHHKKHKKHKK 217
RESULT 15
KZQHL
histidine-rich glycoprotein precursor - Plasmodium lophurae
C:Species: Plasmodium lophurae
C:Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 16-Jul-1999
C:Accession: A22692

R:Ravetch, J.V.; Feder, R.; Pavlovic, A.; Blobel, G.
Nature 312, 616-620, 1984
A:Title: Primary structure and genomic organization of the histidine-rich protein of the
A:Reference number: A22692; MUID:85061618; PMID:6095114
A:Superfamily: pectinesterase pemb
Query Match 61.0%; Score 83; DB 2; Length 391;
Best Local Similarity 68.8%; Pred. No. 0.0047; 4; Indels 0; Gaps 0;
Matches 11; Conservative 1; Mismatches 0; Gaps 0;
QY 5 HKHKKHKKHKKHKK 20
DB 47 HHHHHHHHHHHKK 62
RESULT 13
A29553
histidine-rich protein - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 09-Jun-2000
C:Accession: A29653
R:Lenstra, R.; d'Auriol, L.; Andrieu, B.; Le Bras, J.; Galibert, F.
Biochem. Biophys. Res. Commun. 146, 368-377, 1987
A:Title: Cloning and sequencing of Plasmodium falciparum DNA fragments containing repeti
A:Reference number: A90134; MUID:87270765; PMID:3038111
A:Accession: A29653
A:Molecule type: DNA
A:Residues: 1-82 <LEN>
A:Cross-references: GB:M17028; NID:g160339; PID:g160341
Query Match 60.7%; Score 82.5; DB 2; Length 82;
Best Local Similarity 57.9%; Pred. No. 0.0013;
Matches 11; Conservative 4; Mismatches 3; Indels 1; Gaps 1;
QY 2 HKHKKHKKHKKHKKHKK 20
DB 22 HHHH-HHHHHHHHHHQ 39
RESULT 14
A84696
probable zinc transporter [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: A84696
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: A84696
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-385 <STO>
A:Cross-references: GB:AE002093; NID:g3980394; PIDN:AAC95197.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g29410
A:Map position: 2
Query Match 60.7%; Score 82; DB 1; Length 351;
Best Local Similarity 68.8%; Pred. No. 0.0055;
Matches 11; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 4 HKHKKHKKHKKHKKHKK 19
DB 215 HHHHHHHHHHHHHH 230
Search completed: January 20, 2004, 18:24:44
Job time : 12.1449 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 18:06:29 ; Search time 6.66667 Seconds
(without alignments)
141.080 Million cell updates/sec

Title: US-10-018-103A-5

Perfect score: 136

Sequence: 1 KHKHHKHHKHHKHHKHHK 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89	65.4	139	1 SALA_DROSI	P21749 drosophila
2	89	65.4	142	1 SALA_DROME	P21750 drosophila
3	89	65.4	142	1 SALA_DROO	P21748 drosophila
4	87	64.0	1211	1 BUN2_DROME	Q24523 drosophila
5	85	62.5	657	1 KNOB_PLAFN	P06719 plasmodium
6	83	61.0	428	1 FXB2_MOUSE	Q64733 mus musculus
7	82.5	60.7	82	1 HRP3_PLAFS	P14586 plasmodium
8	82	60.3	351	1 HRPX_PLALO	P04929 plasmodium
9	81.5	59.9	387	1 QIN_AVIS3	P56260 avian sarco
10	81.5	59.9	451	1 FXGB_CHICK	Q90964 gallus gall
11	81.5	59.9	1007	1 PR4B_HUMAN	Q13523 homo sapien
12	81.5	59.9	1007	1 PR4B_MOUSE	Q61136 mus musculus
13	80	58.8	465	1 HNF6_HUMAN	Q9UBC0 homo sapien
14	80	58.8	465	1 HNF6_MOUSE	O08755 mus musculus
15	80	58.8	465	1 HNF6_RAT	P70512 rattus norv
16	79.5	58.5	385	1 SELP_RAT	P25236 rattus norv
17	79.5	58.5	763	1 DYRA_HUMAN	Q13627 homo sapien
18	79.5	58.5	763	1 DYRA_MOUSE	Q61214 mus musculus
19	79.5	58.5	763	1 DYRA_RAT	Q63470 rattus norv
20	79	58.1	143	1 Y970_METJA	Q58380 methanococc
21	79	58.1	143	1 A2AC_DIDMA	P35405 didelphis m
22	78	57.4	193	1 HUNB_DROIK	O46242 drosophila
23	78	57.4	198	1 HUNB_DROCY	O46238 drosophila
24	78	57.4	198	1 HUNB_DRODS	O46240 drosophila
25	78	57.4	198	1 HUNB_DROLI	O46244 drosophila
26	78	57.4	303	1 MOX2_HUMAN	P50222 homo sapien
27	78	57.4	303	1 MOX2_MOUSE	P32443 mus musculus
28	78	57.4	303	1 MOX2_RAT	P39020 rattus norv
29	78	57.4	390	1 PC_DROME	P26017 drosophila
30	78	57.4	430	1 NOR2_RAT	Q63516 rattus norv
31	78	57.4	437	1 ZP12_BRARE	P56224 brachydanio
32	78	57.4	441	1 F033_BRARE	Q90436 brachydanio
33	78	57.4	626	1 NR43_HUMAN	Q92570 homo sapien

34	78	57.4	628	1 NR43_RAT	P51179 rattus norv
35	77.5	57.0	59	1 HPN_HELPY	Q48251 helicobacte
36	77	56.6	420	1 YBE1_SCHPO	O42980 schizosacch
37	76.5	56.2	158	1 HUNB_DRODM	O46248 drosophila
38	76.5	56.2	159	1 HUNB_DROSO	O46258 drosophila
39	76.5	56.2	193	1 HUNB_DROPO	O46252 drosophila
40	76	55.9	149	1 EGGS_SCHMA	P08016 schistosome
41	76	55.9	195	1 HUNB_DRODA	O46262 drosophila
42	76	55.9	196	1 HUNB_DROAA	O46234 drosophila
43	76	55.9	410	1 HNH2_DROME	P10035 drosophila
44	76	55.9	1321	1 PGCN_HUMAN	O14594 homo sapien
45	75.5	55.5	474	1 ANA_DROME	Q26307 drosophila

ALIGNMENTS

RESULT 1					
SALA_DROSI					
ID	SALA_DROSI	STANDARD;	PRT;	139 AA.	
AC	P21749;				
DT	01-MAY-1991 (Rel. 18, Created)				
DT	01-MAY-1991 (Rel. 18, Last sequence update)				
DT	01-FEB-1996 (Rel. 33, Last annotation update)				
DE	Protein spalt-accessory.				
GN	SALA OR SAL.				
OS	Drosophila simulans (Fruit fly).				
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;				
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;				
OC	Ephydroidea; Drosophilidae; Drosophila.				
OX	NCBI_TaxID=7240;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=89315821; PubMed=2568636;				
RA	Reuter D., Schuh R., Jaekle H.;				
RT	"The homeotic gene spalt (sal) evolved during Drosophila speciation.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 86:5483-5486(1989).				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	EMBL; M21227; AAA28877.1; -.				
DR	PIR; B33910; B33910.				
DR	FlyBase; FBgn0012892; Dsml\saala.				
SQ	SEQUENCE 139 AA; 14145 MW; AED5D51561C229CF CRC64;				

Query Match 65.4%; Score 89; DB 1; Length 139;
Best Local Similarity 64.7%; Pred. No. 0.00027;
Matches 11; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY	3	KHKHHKHHKHHKHHK 19
DB	123	EHHEHGHHEHGHHRH 139

RESULT 2					
SALA_DROME					
ID	SALA_DROME	STANDARD;	PRT;	142 AA.	
AC	P21750;				
DT	01-MAY-1991 (Rel. 18, Created)				
DT	01-MAY-1991 (Rel. 18, Last sequence update)				
DT	01-FEB-1995 (Rel. 31, Last annotation update)				
DE	Protein spalt-accessory.				
GN	SALA OR SAL.				
OS	Drosophila melanogaster (Fruit fly).				
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;				
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;				
OC	Ephydroidea; Drosophilidae; Drosophila.				

Wed Jan 21 11:27:52 2004

us-10-018-103a-5.rsp

0;

Matches 11; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

```

OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Frei E., Schuh R., Baumgartner S., Burri M., Noll M., Juergens G.,
RA Seifert E., Nauber U., Jaekle H.;
RT "Molecular characterization of spalt, a homeotic gene required for
RT head and tail development in the Drosophila embryo.";
RL EMBO J. 7:197-204(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89315821; PubMed=2568636;
RA Reuter D., Schuh R., Jaekle H.;
RT "The homeotic gene spalt (sal) evolved during Drosophila speciation.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:5483-5486(1989).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X57474; CAA40712.1; -
CC F01; S00262; S00262.
CC FLYBase; FBgn0003313; sala.
SQ SEQUENCE 142 AA; 14554 MW; 7386D6C8C12E5044 CRC64;

Query Match 65.4%; Score 89; DB 1; Length 142;
Best Local Similarity 64.7%; Pred. No. 0.00028;
Matches 11; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 3 KHKHKKHKKHKKHKKHKKH 19
DB 126 EHHEHGHGHHGHHRH 142

RESULT 3
SALA DROOR
ID SALA DROOR STANDARD; PRT; 142 AA.
AC P21748;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Protein spalt-accessory.
OS SALA OR SAL.
GN Drosophila orena (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7223;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89315821; PubMed=2568636;
RA Reuter D., Schuh R., Jaekle H.;
RT "The homeotic gene spalt (sal) evolved during Drosophila speciation.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:5483-5486(1989).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M21579; AAA28876.1; -
CC F01; C33910;
CC FLYBase; FBgn0012624; Dore\esala.
SQ SEQUENCE 142 AA; 14498 MW; 829A9A586EFEB25F1 CRC64;

Query Match 65.4%; Score 89; DB 1; Length 142;
Best Local Similarity 64.7%; Pred. No. 0.00028;
Matches 11; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

```

0;

Matches 11; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

```

OY 3 KHKHKKHKKHKKHKKHKKH 19
DB 126 EHHEHGHGHHGHHRH 142

RESULT 4
BUN2 DROME
ID BUN2 DROME STANDARD; PRT; 1211 AA.
AC Q24523; Q9VK78; Q9VK79;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Bunched protein, class 2/class 3 isoforms (Shortsighted protein).
GN BUN OR SHS OR CG5463.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM CLASS 2), SUBCELLULAR LOCATION, AND
RP FUNCTION.
RC TISSUE=Eye-antennal disk;
RX MEDLINE=96038094; PubMed=7555710;
RA Treisman J.E., Lai Z.-C., Rubin G.M.;
RT "Shortsighted acts in the decapentaplegic pathway in Drosophila eye
RT development and has homology to a mouse TGF-beta-responsive gene.";
RL Development 121:2835-2845(1995).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS CLASS 2 AND CLASS 3).
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Fessler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Wei M.-H., Ibegwam C.,
RA Hostin D., Houston K.A., Howland T.J., Ke Z., Kennison J.A., Ketchum K.A.,
RA Jalali M., Kalush P., Karpen G.H., Ke Z., Pollard J., Puri V., Reese M.G.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Murphy B., Murphy L., Muszynski D., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pachter J.M., M.G.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J., Yao Q.A.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Zhang G., Zhao Q., Zheng L.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhou S., Zhao Q., Zheng L.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -!- FUNCTION: Probable transcription factor required for peripheral

```

nervous system morphogenesis, eye development and oogenesis. May be required for the transmission of the dpp signal and for a morphogenetic movement of the medulla in the brain that reorients the second optic lobe relative to the first. Plays a role in determining proper dorsal cell fates leading to the formation of the dorsal appendages.

CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.

CC -1- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=3;

CC Comment=Experimental confirmation may be lacking for some isoforms;

CC NamesClass 2;

CC IsoId=Q24523-1; Sequence=Displayed;

CC NamesClass 1;

CC IsoId=Q24522-1; Sequence=External;

CC NamesClass 3;

CC IsoId=Q24523-2; Sequence=VSP_006670;

CC IsoId=Q24522-2; Sequence=VSP_006670;

CC -1- SIMILARITY: BELONGS TO THE TSC-22/DIP/BUN FAMILY.

CC -1- CAUTION: Ref.2 sequence differs from that shown due to erroneous gene model prediction.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; L42512; AAC41608.1; -

CC EMBL; AE003636; AAF53200.1; ALT_SEQ.

CC EMBL; AE003636; AAF53201.1; ALT_SEQ.

CC PIR; T13804; T13804.

CC HSP; P80220; IDIP.

CC FlyBase; FBgn010460; bun.

CC InterPro; IPR000580; TSC-22_Dip_Bun.

CC Pfam; PF01166; TSC22; 1.

CC ProDom; PD007152; TSC-22_Dip_Bun; 1.

CC PROSITE; PS01289; TSC22; 1.

CC Transcription regulation; Nuclear protein; Alternative splicing.

CC DOMAIN 15 31

CC POLY-GLN.

CC FT DOMAIN 76 86

CC ALA-RICH.

CC FT DOMAIN 97 102

CC POLY-GLN.

CC FT DOMAIN 237 241

CC POLY-SER.

CC FT DOMAIN 249 254

CC POLY-GLN.

CC FT DOMAIN 261 265

CC POLY-SER.

CC FT DOMAIN 306 321

CC POLY-HIS.

CC FT DOMAIN 322 328

CC POLY-GLN.

CC FT DOMAIN 348 356

CC POLY-GLY.

CC FT DOMAIN 607 619

CC POLY-GLN.

CC FT DOMAIN 661 668

CC POLY-GLN.

CC FT DOMAIN 743 746

CC POLY-ALA.

CC FT DOMAIN 759 765

CC POLY-GLN.

CC FT DOMAIN 795 801

CC POLY-GLN.

CC FT DOMAIN 817 822

CC POLY-GLN.

CC FT DOMAIN 832 838

CC POLY-ALA.

CC FT DOMAIN 884 891

CC POLY-ALA.

CC FT DOMAIN 927 947

CC GLN-RICH.

CC FT DOMAIN 1001 1005

CC POLY-ALA.

CC FT DOMAIN 1011 1014

CC POLY-ALA.

CC FT DOMAIN 1069 1090

CC LEUCINE-ZIPPER.

CC FT DOMAIN 1194 1201

CC POLY-ALA.

CC FT VARSPLIC 1 109

CC Missing (in isoform Class 3).

CC /FtId=VSP_006670.

CC K -> E (IN REF. 1).

CC MISSING (IN REF. 1).

CC Q -> QQQ (IN REF. 1).

CC QVTSAA -> TS (IN REF. 2).

CC SQ SEQUENCE 1211 AA; 125299 MW; 9925A9159A7051B0 CRC64;

Query Match 64.0%; Score 87; DB 1; Length 1211;

Best Local Similarity 64.7%; Pred. No. 0.0035;

Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 4 HHKHHKHHKHHKHHK 20
||| ||| ||| ||| :
Db 306 HHHHHHHHHHHHHH 322

RESULT 5

KNOB_PLAFN STANDARD; PRT; 657 AA.

AC P06719;

DT 01-JAN-1988 (Rel. 06, Created)

DT 01-JAN-1988 (Rel. 06, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE Knob-associated histidine-rich protein precursor (KAHRP).

GN SD17.

OS Plasmodium falciparum (isolate NF7 / Ghana).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=5842;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=87275836; PubMed=3301325;

RA Triglia T., Stahl H.-D., Crewther P.E., Scanlon D.B., Brown G.V., Anders R.F., Kemp D.J.;

RT The complete sequence of the gene for the knob-associated histidine-rich protein from Plasmodium falciparum.;

RL EMBO J. 6:1413-1419(1987).

CC -1- FUNCTION: KAHRP MIGHT MIMICK HUMAN HISTIDINE-RICH GLYCOPROTEINS TO ANCHOR HOST THROMBOSPONDIN OR A PARASITE ANALOG IN A BINDING COMPLEX WITH THE ENDOTHELIAL CELL RECEPTOR.

CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC SIDE OF THE MEMBRANE OF INFECTED ERYTHROCYTES.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; Y00060; CAAS8268.1; -

CC PIR; A29454; A29454.

CC KW Malaria; Repeat; Signal.

CC FT SIGNAL 1 34

CC POTENTIAL.

CC FT CHAIN 35 657

CC KNOB-ASSOCIATED HISTIDINE-RICH PROTEIN.

CC FT DOMAIN 61 123

CC HIS-RICH.

CC FT CARBOHYD 42 42

CC N-LINKED (GLCNAC...) (POTENTIAL).

CC SQ SEQUENCE 657 AA; 71941 MW; F69E1B9CCF98F946 CRC64;

Query Match 62.5%; Score 85; DB 1; Length 657;

Best Local Similarity 61.1%; Pred. No. 0.0033;

Matches 11; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 KHHKHHKHHKHHKHHK 20
||| ||| ||| ||| :
Db 59 KQHHHHHHHHHHHHH 76

RESULT 6

FXB2_MOUSE STANDARD; PRT; 428 AA.

AC O64733;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Forkhead box protein B2 (transcription factor FKX-4).

GN FOXB2 OR FKX4.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

```

RX MEDLINE=97014266; PubMed=8861101;
RA Kaestner K.H., Schuetz G., Monaghan A.P.;
RT "Expression of the winged helix genes fkh-4 and fkh-5 defines domains
RL in the central nervous system.";
RL Mech. Dev. 55:221-230(1996).
RN [2]
RP SEQUENCE OF 4-114 FROM N.A.
RC STRAIN=129;
RX MEDLINE=93361500; PubMed=7689224;
RA Kaestner K.H., Lee K.H., Schloendorff J., Hiemisch H.,
RA Monaghan A.P., Schuetz G.;
RT "Six members of the mouse forkhead gene family are developmentally
RL regulated.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:7628-7631(1993).
RC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING EMBRYOGENESIS.
CC -1- SIMILARITY: Contains 1 fork-head domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X92591; CAA63335.1; -.
DR EMBL; X71942; CAA50744.1; -.
DR PIR; D47746; D47746.
DR HSP; O63245; 2HPH.
DR TRANSFAC; T02442; -.
DR MGD; MGI:1347468; Foxb2.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS00399; FORK_HEAD_3; 1.
DR DNA-binding; Nuclear protein; Transcription regulation.
KW DNA_BIND 12 103
FT DOMAIN 129 153
FT POLY-HIS.
FT DOMAIN 156 162
FT POLY-HIS.
FT DOMAIN 163 172
FT POLY-PRO.
FT DOMAIN 217 231
FT POLY-ALA.
FT DOMAIN 249 258
FT POLY-ALA.
FT DOMAIN 321 330
FT POLY-ALA.
FT DOMAIN 396 399
FT POLY-ALA.
SQ SEQUENCE 428 AA; 45170 MW; DB8A8EFD1E94AB10 CRC64;

Query Match 61.0%; Score 83; DB 1; Length 428;
Best Local Similarity 68.8%; Pred. No. 0.0036;
Matches 11; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 HKHHKHHKHHKHHKHH 17
DB 137 HPHHPHHHHHHHHH 152

RESULT 7
HRP3_PLAFS STANDARD; PRT; 82 AA.
AC P14586;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-MAY-1991 (Rel. 18, Last annotation update)
DE Histidine-rich protein.
OS Plasmodium falciparum (isolate fcmu7 / Senegal).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5845;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=87270765; PubMed=3038111;
RA Lenstra R., D'Auriol L., Andrieu B., le Bras J., Galibert F.;
RT "Cloning and sequencing of Plasmodium falciparum DNA fragments
RT containing repetitive regions potentially coding for histidine-rich
RL proteins: identification of two overlapping reading frames.";
RL Biochem. Biophys. Res. Commun. 146:368-377(1987).
CC -1- MISCELLANEOUS: THIS PROTEIN IS CODED ON THE REVERSE STRAND OF AN
CC ASPARTIC ACID-RICH PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M17028; AAA29619.1; -.
DR PIR; A29653; A29653.
KW Malaria.
SQ SEQUENCE 82 AA; 10578 MW; 7CE2EA69F2FC1E8C CRC64;

Query Match 60.7%; Score 82.5; DB 1; Length 82;
Best Local Similarity 57.9%; Pred. No. 0.00087;
Matches 11; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

QY 2 HKHHKHHKHHKHHKHHK 20
DB 22 HRRHH-HRRHHHHRRHHHQ 39

RESULT 8
HRP3_PLALO STANDARD; PRT; 351 AA.
AC P04929;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histidine-rich glycoprotein precursor.
OS Plasmodium lophurae.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5853;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=85061618; PubMed=6095114;
RA Ravetch J.V., Feder R., Pavlovic A., Blobel G.;
RT "Primary structure and genomic organization of the histidine-rich
RT protein of the malaria parasite Plasmodium lophurae.";
RL Nature 312:616-620(1984).
CC -1- MISCELLANEOUS: IN THE INTRAERYTHROCYTIC STAGES OF DEVELOPMENT OF
CC P. LOPHURAE IN DUCKS, THERE IS A SYNTHESIS OF A MAJOR PROTEIN THAT
CC ACCUMULATES TO COMPRISE AT LEAST 50% OF THE CELLULAR MASS: THE
CC HISTIDINE RICH PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X01469; CAA25698.1; -.
DR PIR; A22692; KGZQHL.
KW Malaria; Repeat; Glycoprotein; Signal.
FT SIGNAL 1 23
FT PROPEP 24 47
FT CHAIN 48 351
FT CARBOHYD 40 40
FT DOMAIN 59 90
FT REPEAT 59 74
FT REPEAT 75 90
FT DOMAIN 91 123
DE HISTIDINE-RICH GLYCOPROTEIN.
DE N-LINKED (GLCNAC...) (PROBABLE).
DE 2 X 16 AA TANDEM REPEATS.
DE 16-1.
DE REPEAT 75 90
DE 2 X 17 AA TANDEM REPEATS.

```



```
FT REPEAT 91 107 17-1.
FT REPEAT 108 123 17-2.
FT DOMAIN 124 153 2 X 15 AA TANDEM REPEATS.
FT REPEAT 124 138 15-1.
FT REPEAT 139 153 15-2.
FT DOMAIN 173 351 18 X 10 AA TANDEM REPEATS.
SQ SEQUENCE 351 AA; 44032 MW; D19A48D47D890453 CRC64;

Query Match 60.3%; Score 82; DB 1; Length 351;
Best Local Similarity 68.8%; Pred. No. 0.0039;
Matches 11; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 4 HKHKKHKKHKKHKKH 19
||| ||| ||| ||| |||
Db 215 HHHHHHHGHHHHHHH 230

RESULT 9
QIN_AVIS3
ID_QIN_AVIS3 STANDARD; PRT; 387 AA.
AC P56260;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transforming protein QIN (Oncogene QIN).
GN V-QIN.
OS Avian sarcoma virus (strain 31) (ASV31).
OC Viruses; Retroviridae; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=35270;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93281605; PubMed=8099441;
RA Li J., Vogt P.K.;
RT "The retroviral oncogene qin belongs to the transcription factor
RT family that includes the homotetic gene fork head.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:4490-4494(1993).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-QIN
CC POLYPROTEIN.
CC -1- SIMILARITY: Contains 1 fork-head domain.
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L10719; -; NOT ANNOTATED_CDS.
CC HSP; Q63245; 2HFH.
CC TRANSFAC; T01832; -.
CC InterPro; IPR001766; TF_Fork_head.
CC Pfam; PF00250; Fork_head; 1.
CC PRINTS; PR00053; FORKHEAD.
CC ProDom; PD000425; TF_Fork_head; 1.
CC SMART; SM00339; FH; 1.
CC PROSITE; PS00657; FORK_HEAD_1; 1.
CC PROSITE; PS00658; FORK_HEAD_2; 1.
CC PROSITE; PS50039; FORK_HEAD_3; 1.
CC DNA-binding; Nuclear protein; Oncogene.
CC DOMAIN 42 54 POLY-PRO.
CC FT DOMAIN 55 58 POLY-PRO.
CC FT DOMAIN 64 67 POLY-ALA.
CC FT DOMAIN 101 106 POLY-ALA.
CC FT DNA_BIND 142 233 FORK-HEAD.
CC SEQUENCE 387 AA; 42283 MW; FEA902F50FF42F9 CRC64;

Query Match 59.9%; Score 81.5; DB 1; Length 387;
Best Local Similarity 63.2%; Pred. No. 0.0049;
Matches 12; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

Qy 2 HKHH-KHHKKHKKHKKH 19
```

```
Db 36 HSHSHSHPHHHHHHHH 54

RESULT 10
FXGB_CHICK
ID_FXGB_CHICK STANDARD; PRT; 451 AA.
AC Q90964;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Forkhead box protein G1B (Forkhead-related protein FXH1)
DE (Transcription factor BF-1) (Brain factor 1) (BFI) (CBF-1) (Proto-
DE oncogene C-QIN) (N-62-5) (CEQ 3-1).
GN FOXG1B OR FXH1 OR QIN.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95132616; PubMed=7831308;
RA Chang H.W., Li J., Kretschmar D., Vogt P.K.;
RT "Avian cellular homolog of the qin oncogene.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:447-451(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=White leghorn; TISSUE=Retina;
RX MEDLINE=96338226; PubMed=8757134;
RA Yuasa J., Hirano S., Yamagata M., Noda M.;
RT "Visual projection map specified by topographic expression of
RT transcription factors in the retina.";
RL Nature 382:632-635(1996).
CC -1- FUNCTION: MAY DETERMINE THE NASOTEMPORAL AXIS OF THE RETINA, AND
CC CONSEQUENTLY SPECIFY THE TOPOGRAPHICAL PROJECTION OF THE RETINAL
CC GANGLION-CELL AXONS TO THE TECTUM BY CONTROLLING EXPRESSION OF
CC THEIR TARGET GENES.
CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
CC -1- TISSUE SPECIFICITY: RETINA AND BRAIN.
CC -1- DEVELOPMENTAL STAGE: CAN BE DETECTED IN REGIONS INCLUDING
CC PRIMITIVE RETINA AND NEUROEPITHELIUM BY EMBRYONIC DAY 2 (E2). AT
CC E3, EXPRESSED IN THE NASAL RETINA AND PIGMENT EPITHELIUM AS WELL
CC AS IN THE TELENCEPHALON, AND AT E7 IS EXPRESSED IN RETINAL
CC GANGLION CELLS. LEVELS BEGIN TO DECLINE FROM E4 AND ALMOST
CC DISAPPEAR BY E10.
CC -1- SIMILARITY: Contains 1 fork-head domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L36814; AAA66954.1; -.
CC EMBL; U47275; AAB08466.1; -.
CC FIR; A55909; A55909.
CC HSP; Q63245; 2HFH.
CC TRANSFAC; T01833; -.
CC InterPro; IPR001766; TF_Fork_head.
CC Pfam; PF00250; Fork_head; 1.
CC PRINTS; PR00053; FORKHEAD.
CC ProDom; PD000425; TF_Fork_head; 1.
CC SMART; SM00339; FH; 1.
CC PROSITE; PS00657; FORK_HEAD_1; 1.
CC PROSITE; PS00658; FORK_HEAD_2; 1.
CC PROSITE; PS50039; FORK_HEAD_3; 1.
CC Transcription regulation; DNA-binding; Nuclear protein;
CC Developmental protein; Proto-oncogene.
CC DOMAIN 45 54 POLY-HIS.
CC FT DOMAIN 55 58 POLY-PRO.
```


RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli P., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Ronaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Koncaiki S.,
 RA Hayashizaki Y.;
 RA "Functional annotation of a full-length mouse cDNA collection.";
 RT Nature 409:685-690(2001).
 RN [3]
 RP SEQUENCE OF 185-1007 FROM N.A.
 RX MEDLINE=98369054; PubMed=9701556;
 RA Tate P., Lee M., Tweedie S., Skarnes W., Bickmore W.;
 RT "Capturing novel mouse genes encoding chromosomal and other nuclear
 RT proteins.";
 RL J. Cell Sci. 111:2575-2595(1998).
 RN [4]
 RP SEQUENCE OF 512-1007 FROM N.A.
 RX MEDLINE=97250302; PubMed=9102632;
 RA Gross T., Lutzelberger M., Wiegmann H., Klingenhoff A., Shenoy S.,
 RA Kaeuffer N.P.;
 RT "Functional analysis of the fission yeast Prp4 protein kinase
 RT involved in pre-mRNA splicing and isolation of a putative mammalian
 RT homologue.";
 RL Nucleic Acids Res. 25:1028-1035(1997).
 CC -!- FUNCTION: Has a role in pre-mRNA splicing. Phosphorylates SF2/ASF.
 CC -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
 CC -!- SUBUNIT: Interacts with Clk1 C-terminus.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- PTM: Phosphorylated by Clk1.
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AF283466; AM19102.1; -;
 DR EMBL; AK020579; BAB32137.1; -;
 DR EMBL; AK021274; BAB32358.1; -;
 DR EMBL; AF033663; AAC32042.1; -;
 DR EMBL; U48737; AAB03269.1; -;
 DR HSSP; P24941; 1AQL.
 DR MCD; MGI:109584; Prpf4b.
 DR GO; GO:0005694; C:chromosome; IDA.
 DR InterPro; IPR000719; Prot kinase.
 DR InterPro; IPR002290; Ser Thr pkinase.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR SMART; SM00220; S_TKc; 1.
 DR SMART; SM00219; TyrKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 KW mRNA processing; mRNA splicing; Transferase;
 KW Serine/threonine-protein kinase; ATP-binding; Phosphorylation;
 KW Nuclear protein.
 FT DOMAIN 41 79 HIS-RICH.
 FT DOMAIN 40 497 ARG/LYS-RICH.
 FT DOMAIN 687 1006 PROTEIN_KINASE.
 FT NP_BIND 693 701 ATP (BY SIMILARITY).
 FT BINDING 717 717 ATP (BY SIMILARITY).
 FT

FT ACT_SITE 815 815 BY SIMILARITY.
 FT CONFLICT 185 187 SKS -> IPG (IN REF. 3).
 FT CONFLICT 223 223 K -> I (IN REF. 3).
 FT CONFLICT 633 633 F -> L (IN REF. 4).
 SQ SEQUENCE 1007 AA; 116947 MW; 18E6C3C43BE7AB4C CRC64;
 Query Match 59.9%; Score 81.5; DB 1; Length 1007;
 Best Local Similarity 54.5%; Pred. No. 0.012;
 Matches 18; Conservative 0; Mismatches 2; Indels 13; Gaps 3;
 QY 1 KHKKH---KHKKH-----HKH-KHKKHKK 20
 DB 48 KHKKSKKKKKHKKHSSSEDRDKKKHKKHKKHKK 80
 RESULT 13
 HNF6 HUMAN
 ID HNF6 HUMAN STANDARD; PRT; 465 AA.
 AC Q9UBC0; Q99744; Q9UMR6;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hepatocyte nuclear factor 6 (HNF-6) (One cut domain family member 1).
 GN ONECUT1 OR HNF6A OR HNF6.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Furuta H., Horikawa Y., Wang Y.-Q., Bell G.I.;
 RT "Isolation and characterization of the human hepatocyte nuclear factor
 RT 6 gene.";
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Furuta H., Wang Y.-Q., Bell G.I.;
 RT "The sequence of human mRNA for the hepatocyte nuclear factor-6
 RT alpha.";
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-368 FROM N.A., AND VARIANT ALA-75.
 RX MEDLINE=99420592; PubMed=10491763;
 RA Moeller A.M., Ek J., Durvieux S.M., Urhammer S.A., Clausen J.O.,
 RA Eiberg H., Hansen T., Rousseau G.G., Lemaigre F.P., Pedersen O.;
 RT "Hepatocyte nuclear factor-6: associations between genetic variability
 RT and type II diabetes and between genetic variability and estimates of
 RT insulin secretion.";
 RL Diabetologia 42:1011-1016(1999).
 RN [4]
 RP SEQUENCE OF 174-465 FROM N.A.
 RA Samadani U., Costa R.H.;
 RT "Yeast one-hybrid cloning of the partial human cDNA for hepatocyte
 RT nuclear factor 6.";
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: TRANSCRIPTIONAL ACTIVATOR. BINDS THE CONSENSUS SEQUENCE
 CC 5'-DHATTGATTTWD-3' ON A VARIETY OF GENE PROMOTERS SUCH AS THOSE
 CC OF HNF3B AND TTR. IMPORTANT FOR LIVER GENES TRANSCRIPTION.
 CC -!- SUBUNIT: BINDS DNA AS A MONOMER (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN LIVER; LOWER EXPRESSION IN
 CC TESTIS AND SKIN.
 CC -!- SIMILARITY: Contains 1 CUT domain.
 CC -!- SIMILARITY: BELONGS TO THE CUT HOMEBOX FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC

```

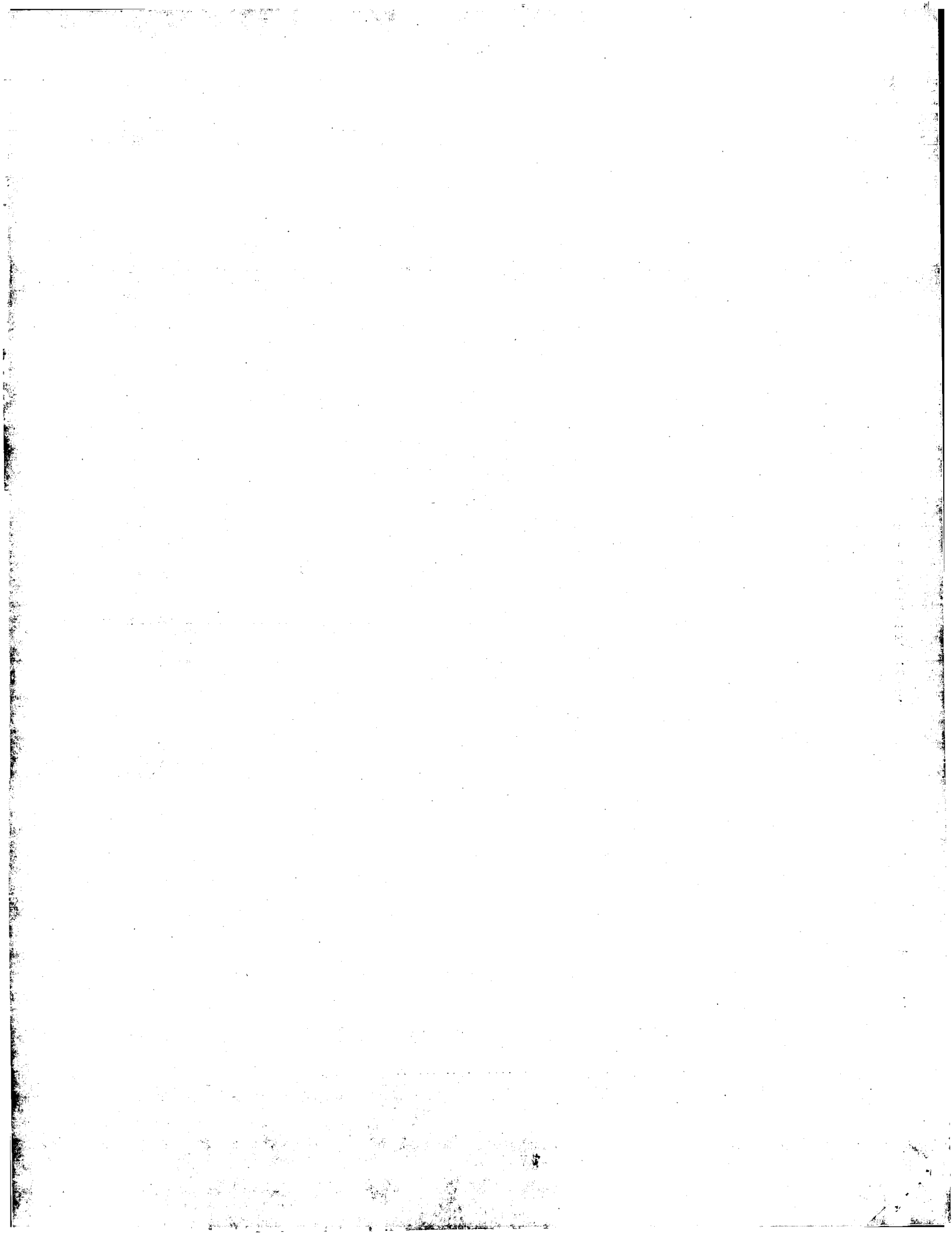
CC EMBL; AF035581; AAD02033.1; --
CC EMBL; AF035580; AAD02033.1; JOINED.
CC EMBL; U96173; AAD00826.1; --
CC EMBL; Y17739; CAB50769.1; --
CC EMBL; U77975; AAB61705.1; --
CC TRANSFAC; T03286; --
CC Genew; HGNC:8138; ONECUT1.
CC MIM; 604164; -- C:nucleus; NAS.
CC GO; GO:0005634; C:nucleus; NAS.
CC GO; GO:0016563; F:transcriptional activator activity; NAS.
CC GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
CC InterPro; IPR007108; Cut.homo.
CC InterPro; IPR001356; Homeobox.
CC Pfam; PF02376; CUT; 1.
CC Pfam; PF00046; homeobox; 1.
CC ProDom; PD000010; Homeobox; 1.
CC SMART; SM00389; HOX; 1.
CC PROSITE; PS00027; HOMEBOX 1; FALSE_NEG.
CC PROSITE; PS00071; HOMEBOX 2; 1.
CC Transcription regulation; Homeobox; DNA-binding; Nuclear protein;
KW Activator; Polymorphism.
FT DNA BIND 283 369
FT DNA BIND 385 444
FT DOMAIN 69 72
FT DOMAIN 124 138
FT DOMAIN 455 460
FT VARIANT 75 75
FT CONFLICT 220 220
FT CONFLICT 284 284
FT CONFLICT 288 288
FT CONFLICT 318 318
FT CONFLICT 386 386
FT SEQUENCE 465 AA; 51023 MW; F47F78957A65CFC2 CRC64;
Query Match 58.8%; Score 80; DB 1; Length 465;
Best Local Similarity 55.0%; Pred. No. 0.0085;
Matches 11; Conservative 1; Mismatches 8; Indels 0; Gaps 0;
QY 1 KHKHKHKHKHKHKHKHK 20
DB 121 KFPFHHPHHHPHHHPHQR 140

RESULT 14
HNF6_MOUSE
ID HNF6_MOUSE STANDARD; PRT; 465 AA.
AC O08755;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hepatocyte nuclear factor 6 (HNF-6) (One cut domain family member 1).
GN ONECUT1 OR HNF6A OR HNF6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=98104231; PubMed=9441664;
RA Rausa F., Samadani U., Ye H., Lim L., Fletcher C.F., Jenkins N.A.,
RA Copeland N.G., Costa R.H.;
RT "The cut-homeodomain transcriptional activator HNF-6 is coexpressed
RT with its target gene HNF-3 beta in the developing murine liver and
RT pancreas."
RL Dev. Biol. 192:228-246(1997).
CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR. BINDS THE CONSENSUS SEQUENCE
CC 5'-DHATTGATTTWD-3', ON A VARIETY OF GENE PROMOTERS SUCH AS THOSE
CC OF HNF3B AND TTR. IMPORTANT FOR LIVER GENES TRANSCRIPTION (BY
CC SIMILARITY).
CC -1- SUBUNIT: BINDS DNA AS A MONOMER (BY SIMILARITY).

-1- SUBCELLULAR LOCATION: Nuclear.
-1- SIMILARITY: Contains 1 CUT domain.
-1- SIMILARITY: BELONGS TO THE CUT HOMEBOX FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----
EMBL; U95945; AAB53863.1; --
TRANSFAC; T03296; --
MGD; MGI:1196423; Onecut1.
GO; GO:0003700; F:transcription factor activity; IDA.
GO; GO:0030154; P:cell differentiation; IMP.
GO; GO:0007492; P:endoderm development; IMP.
GO; GO:0006006; P:glucose metabolism; IMP.
InterPro; IPR007108; Cut.homo.
InterPro; IPR001350; Hmoeo CUT.
InterPro; IPR001356; Homeobox.
Pfam; PF02376; CUT; 1.
Pfam; PF00046; homeobox; 1.
ProDom; PD000010; Homeobox; 1.
SMART; SM00389; HOX; 1.
PROSITE; PS00027; HOMEBOX 1; FALSE_NEG.
PROSITE; PS00071; HOMEBOX 2; 1.
Transcription regulation; Homeobox; DNA-binding; Nuclear protein;
KW Activator.
FT DNA BIND 283 369
FT DNA BIND 385 444
FT DOMAIN 69 72
FT DOMAIN 124 138
FT DOMAIN 450 460
FT SEQUENCE 465 AA; 50952 MW; A770D27DD5AAC896 CRC64;
Query Match 58.8%; Score 80; DB 1; Length 465;
Best Local Similarity 55.0%; Pred. No. 0.0085;
Matches 11; Conservative 1; Mismatches 8; Indels 0; Gaps 0;
QY 1 KHKHKHKHKHKHKHKHK 20
DB 121 KFPFHHPHHHPHHHPHQR 140

RESULT 15
HNF6_RAT
ID HNF6_RAT STANDARD; PRT; 465 AA.
AC P70512; O08755;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hepatocyte nuclear factor 6 (HNF-6) (One cut domain family member 1).
GN ONECUT1 OR HNF6A OR HNF6.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM ALPHA).
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=96382488; PubMed=8790352;
RA Lemaigre F.P., Durvieux S.M., Truong O., Lannoy V.J., Hauan J.J.,
RA Rousseau G.G.;
RT "Hepatocyte nuclear factor 6, a transcription factor that contains a
RT novel type of homeodomain and a single cut domain."
RL Proc. Natl. Acad. Sci. U.S.A. 93:9460-9464(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM BETA).
RC STRAIN=Wistar;
RX MEDLINE=98256275; PubMed=9593691;
RA Lannoy V.J., Bueglin T.R., Rousseau G.G., Lemaigre F.P.;

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 18:20:39 ; Search time 24.6377 Seconds
(without alignments)
209.478 Million cell updates/sec

Title: US-10-018-103A-5
Perfect score: 136
Sequence: 1 KHGHKHHKHHKHHKHHK

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL_23:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_rvirus:*
 - 16: sp_bacteriap:*
 - 17: sp_archesp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	97	71.3	186	5 Q9V3W5	Q9v3w5 drosophila
2	97	71.3	3384	5 Q8II42	Q8ii42 plasmodium
3	94	69.1	1413	5 Q9VJ38	Q9vj38 drosophila
4	94	69.1	1700	5 Q9VJ39	Q9vj39 drosophila
5	92	67.6	326	16 Q9I447	Q9i447 pseudomonas
6	92	67.6	404	13 Q8UVJ4	Q8uvj4 petromyzon
7	91	66.9	684	10 Q9SG87	Q9sg87 arabidopsis
8	91	66.9	782	5 Q8TU29	Q8tu29 drosophila
9	91	66.9	825	5 Q8T3U6	Q8t3u6 drosophila
10	91	66.9	1291	5 Q77261	Q77261 drosophila
11	91	66.9	1300	5 Q9WE0	Q9we0 drosophila
12	90	66.2	233	5 Q8MP30	Q8mp30 dictyosteli
13	90	66.2	407	10 Q9ZQA3	Q9zqa3 arabidopsis
14	89.5	65.8	964	5 Q9VKD8	Q9vkd8 drosophila
15	89.5	65.8	970	5 Q95TU2	Q95tu2 drosophila
16	89	65.4	142	5 Q9VKH1	Q9vkh1 drosophila

17	89	65.4	145	13 Q90ZQ7	Q90zq7 oryzias lat
18	89	65.4	325	4 Q9BSN0	Q9bsn0 homo sapien
19	89	65.4	449	11 Q8C4J8	Q8c4j8 mus musculu
20	89	65.4	451	4 Q969F2	Q969f2 homo sapien
21	89	65.4	461	11 Q9D7U9	Q9d7u9 mus musculu
22	89	65.4	461	11 Q9IY45	Q9iy45 mus musculu
23	89	65.4	461	11 Q8VE28	Q8ve28 mus musculu
24	89	65.4	1122	5 Q8SXD4	Q8sxd4 drosophila
25	89	65.4	1281	13 Q9I8L4	Q9i8l4 brachydanio
26	87	64.0	89	5 Q8MP90	Q8mp90 dictyosteli
27	87	64.0	1040	13 Q90840	Q90840 gallus gall
28	87	64.0	1057	11 Q9EQM4	Q9eqm4 mus musculu
29	86.5	63.6	503	5 Q9I7K2	Q9i7k2 drosophila
30	86.5	63.6	652	5 Q95PE9	Q95pe9 drosophila
31	86	63.2	1147	5 Q95PH8	Q95ph8 dictyosteli
32	86	63.2	1365	5 Q9VJ79	Q9vj79 drosophila
33	86	63.2	1544	5 Q8MQW0	Q8mqw0 drosophila
34	85	62.5	1318	5 Q9SPH4	Q9sph4 dictyosteli
35	84	61.8	70	4 Q9BSX0	Q9bsx0 homo sapien
36	84	61.8	148	4 Q9NZ81	Q9ntz8 homo sapien
37	84	61.8	148	4 Q9NTQ8	Q9ntq8 homo sapien
38	84	61.8	868	5 Q9VBH3	Q9vbh3 drosophila
39	83.5	61.4	769	5 Q8MRL5	Q8mrl5 drosophila
40	83.5	61.4	954	5 Q8ILV8	Q8ilv8 plasmodium
41	83.5	61.4	1033	10 Q9XEX5	Q9xex5 chlamydomon
42	83.5	61.4	1059	5 Q9W493	Q9w493 gossypium h
43	83	61.0	314	10 Q9SPJ7	Q9spj7 gossypium h
44	83	61.0	391	10 Q23038	Q23038 arabidopsis
45	83	61.0	765	11 Q8C145	Q8c145 mus musculu

ALIGNMENTS

RESULT 1

Q9V3W5	PRELIMINARY;	PRT;	186 AA.
ID	Q9V3W5		
AC	Q9V3W5;		
DT	01-MAY-2000 (Tremblrel. 13, Created)		
DT	01-OCT-2002 (Tremblrel. 22, Last sequence update)		
DT	01-MAR-2003 (Tremblrel. 23, Last annotation update)		
DE	CG14052 protein.		
GN	CG14052.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;		
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Berkeley;		
RX	MEDLINE=20196006; PubMed=10731132;		
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.P.,		
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,		
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,		
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,		
RA	Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Miklos G.L.G.,		
RA	Balieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,		
RA	Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,		
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,		
RA	Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,		
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,		
RA	de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,		
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,		
RA	Dubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,		
RA	Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,		
RA	Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,		
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,		
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,		
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,		
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,		

Wed Jan 21 11:27:52 2004

```

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E.C., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "the genome sequence of Drosophila melanogaster.";
RT Science 287:2185-2195(2000).
[2]
SEQUENCE FROM N.A.
RA Celnik S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne S.E., Nianatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattel B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradscky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celnik S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.B.;
RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[4]
SEQUENCE FROM N.A.
RA Adams M.D., Celnik S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[5]
SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB003423; AAF45754.2; -.
DR FlyBase; FBgn0029606; CG14052.
SQ SEQUENCE 186 AA; 21848 MW; 9BF945B6AD4EFA9D CRC64;

Query Match 71.3%; Score 97; DB 5; Length 186;
Best Local Similarity 68.4%; Pred. No. 1.8e-05;
Matches 13; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 KKHKKHKKHKKHKKHKKH 19
DB 118 KKHKKHKKHKKHKKHKKH 136

RESULT 2
Q81L42 PRELIMINARY; PRT; 3384 AA.
AC Q81L42;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

us-10-018-103a-5.rspt

DE Hypothetical protein.
GN PF14_0407.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RE MEDLINE=22255705; PubMed=12368864;
RX Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014822; AAN37020.1; -.
KW Hypothetical protein.
SQ SEQUENCE 3384 AA; 404576 MW; 340A187BCAEBAG6FD CRC64;

Query Match 71.3%; Score 97; DB 5; Length 3384;
Best Local Similarity 63.2%; Pred. No. 0.00025;
Matches 12; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 KKHKKHKKHKKHKKHKKH 19
DB 647 EHNHNHNHNHNHNHNHNH 665

RESULT 3
Q9VJU8 PRELIMINARY; PRT; 1413 AA.
AC Q9VJU8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE C34668 protein (LD30602P).
GN C331738 OR CG4668 OR CG13261.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RE MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Posler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

```


RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarly C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M., Reese M.G.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stappleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003652; AAF53547.2; -
DR EMBL; AY094813; AM11166.1; -
DR FlyBase; FBgn0051738; CG31738.
DR InterPro; IPR003962; FNIII subd.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR001865; Ribosomal_S2.
DR Pfam; PF00041; fn3; 9
DR PRINTS; PR00014; FNTYEIII.
DR SMART; SM00060; FN3; 9.
DR PROSITE; PS00962; RIBOSOMAL_S2_1; 1.
KW Repeat.
SQ SEQUENCE 1413 AA; 153288 MW; 93390C986EB40B03 CRC64;
Query Match 69.1%; Score 94; DB 5; Length 1413;
Best Local Similarity 63.2%; Pred. No. 0.00027;
Matches 12; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 1 KKHKKKKKKKKKKKKKK 19
Db 1321 QHQHHHHHHHHHHHHHH 1339
RESULT 4
ID Q9VJ9 PRELIMINARY; PRT; 1700 AA.
AC Q9VJ9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG31738 protein.
GN CG31738 OR CG14668 OR CG13261.
OS *Drosophila melanogaster* (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; *Drosophila*.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballwey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler K., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimball B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarly C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stappleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Dou L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferrera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stappleton M., Strong R., Svirkas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of *Drosophila melanogaster* genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Klomp B., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of *Drosophila melanogaster* genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003652; AAF53546.3; -
DR FlyBase; FBgn0051738; CG31738.
DR InterPro; IPR003962; FNIII subd.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR001865; Ribosomal_S2.

Wed Jan 21 11:27:52 2004

us-10-018-103a-5.rpt

```

DR Pfam; PF00041; fn3; 9;
DR PRINTS; PR00014; FNTPEIIL.
DR SMART; SM00060; FN3; 9;
DR PROSITE; PS00962; RIBOSOMAL_S2_1; 1.
KW Repeat.
SQ SEQUENCE 1700 AA; 183716 MW; 4882318D37AAD052 CRC64;

Query Match
Best Local Similarity 69.1%; Score 94; DB 5; Length 1700;
Matches 12; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 KHKKHKKHKKHKKHKKHKK 19
Db 1608 QHQHHMHSHMHHAHPH 1626

RESULT 5
ID Q91447 PRELIMINARY; PRT; 326 AA.
AC Q91447;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Probable metal transporter.
GN PA1297.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warriner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Golltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.A., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.B.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen";
RL Nature 406:959-964 (2000).
DR EMBL; AE004559; AAG04686.1; -.
DR InterPro; IPR002524; Cation efflux.
DR Pfam; PF01545; Cation efflux; 1.
DR TIGRfam; TIGR01297; CDF; 1.
KW Complete proteome.
SQ SEQUENCE 326 AA; 35772 MW; F948E09ACC2DB92A CRC64;

Query Match
Best Local Similarity 67.6%; Score 92; DB 16; Length 326;
Matches 12; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 HKHKKHKKHKKHKKHKKHKK 20
Db 163 HHHHHHHHHHHHHHHAHR 181

RESULT 6
ID Q8UVJ4 PRELIMINARY; PRT; 404 AA.
AC Q8UVJ4;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hox1w (Fragment).
GN Hox1w.
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
OX NCBI_TaxID=7757;
RN [1]
RP SEQUENCE FROM N.A.

Force A., Amores A., Postlethwait J.;
"Fox cluster organization in the jawless vertebrate, Petromyzon
marinus, and the evolution of the vertebrate Hox clusters.";
Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
-1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
EMBL; AF434665; AAL61641.1; -.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; Hox; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
FT NON_TER 1 404
FT SEQUENCE 404 AA; 41691 MW; A01537E167772552 CRC64;

Query Match
Best Local Similarity 67.6%; Score 92; DB 13; Length 404;
Matches 12; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 HKHKKHKKHKKHKKHKKHKK 20
Db 70 HHHHHHHHHHHHHHHHQ 88

RESULT 7
ID Q9SG87 PRELIMINARY; PRT; 684 AA.
AC Q9SG87;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Putative RING zinc finger protein.
GN T7M13.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X., Kaul S., Town C.D., Benito M.-I., Creasy T.H., Haas B.,
RA Renning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RA "Arabidopsis thaliana chromosome III BAC T7M13 genomic sequence.";
RT Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
-1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC EMBL; AC011708; AAF19568.1; -.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00097; zf-C3HC4; 1.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00089; ZF_RING_2; 1.
KW Metal-binding; zinc; zinc-finger.
SQ SEQUENCE 684 AA; 76659 MW; 946203A42A7399B1 CRC64;

Query Match
Best Local Similarity 66.7%; Score 91; DB 10; Length 684;
Matches 12; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 HKHKKHKKHKKHKKHKKHKK 19
Db 520 HHHHHHHHHHHHHHHHHH 537

RESULT 8
ID Q8T029 PRELIMINARY; PRT; 782 AA.
AC Q8T029;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)

```


Wed Jan 21 11:27:52 2004

```

RA Fosler C., Gabriellan A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Markulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Strong R., Sun E.,
RA Spier E., Spradling A.C., Stapleton M., Turner E., Wang A.H., Wang X.,
RA Svirskas R., Tector C., Turner R., Weinstock G.M., Weissenbach J.,
RA Wang Z.-Y., Wassarman D.A., Weinstein K.C., Wu D., Yang S., Yao Q.A.,
RA Williams S.M., Woodage T., Worley K.C., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster";
RT Science 287:2185-2195(2000).
RL -1- SIMILARITY: CONTAINS 1 SET DOMAIN.
CC EMBL; AB003418; AAF4537.1; --
DR FLYBase; FBgn0025639; EG:115C2.10.
DR InterPro; IPR001214; SET.
DR Pfam; PF00856; SET; 1.
DR SMART; SM00317; SET; 1.
DR PROSITE; PS00280; SET; 1.
SQ SEQUENCE 1300 AA; 137524 MW; 7186D0198CC16926 CRC64;

Query Match 66.9%; Score 91; DB 5; Length 1300;
Best Local Similarity 66.7%; Pred. No. 0.00059;
Matches 12; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 HKHKKHKKHKKHKKHKKH 19
Db 539 HQHFFHHHHHHHHHHH 556

RESULT 12
Q8MF30 PRELIMINARY; PRT; 233 AA.
AC Q8MF30;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA Lehmann K., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and Analysis of Chromosome 2 of Dictyostelium."
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC123513; AAM44363.1; --
KW Hypothetical protein.
SQ SEQUENCE 233 AA; 28937 MW; AD4C7A2F86BB14B9 CRC64;

Query Match 66.2%; Score 90; DB 5; Length 233;
Best Local Similarity 66.7%; Pred. No. 0.00017;
Matches 12; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 HKHKKHKKHKKHKKHKKH 19
Db 71 HHHHHHHHHHHHHHHHHHH 88

RESULT 13
Q9ZQA3 PRELIMINARY; PRT; 407 AA.
AC Q9ZQA3;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Putative pectinesterase.
GN AT3G36710.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Eumetazoa; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eumetazoa II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buel C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Niernan W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RA "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana";
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006282; AAD20147.1; --
DR InterPro; IPR000070; Pectinesterase.
DR Pfam; PF01095; Pectinesterase; 1.
DR PROSITE; PS00503; PECTINESTERASE 2; 1.
SQ SEQUENCE 407 AA; 44973 MW; DBA7B9E166D0B42 CRC64;

Query Match 66.2%; Score 90; DB 10; Length 407;
Best Local Similarity 66.7%; Pred. No. 0.00027;
Matches 12; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 HKHKKHKKHKKHKKHKKH 19
Db 50 HKHHHHHHHHHHHHHHHHH 67

RESULT 14
Q9VKD8 PRELIMINARY; PRT; 964 AA.
AC Q9VKD8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG6866 protein.
GN CG6866.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

```


Wed Jan 21 11:27:52 2004

us-10-018-103a-5.rspt

```

RN SEQUENCE FROM N.A.
RP Celnikier S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Anantides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferriera S., Frise E., Galle R.F., Gaig N.S., George R.A.,
RA Gonzales M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B., Scheeler F.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirska R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Miara S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaniker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celnikier S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celnikier S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY058525; AAL13754.1; -.
DR EMBL; AE003634; AAF53138.2; -.
DR FlyBase; FBgn0032388; CG6686.
DR InterPro; IPR005011; SART_1.
DR Pfam; PF03343; SART-1; 1.
SQ SEQUENCE 970 AA; 112575 MW; A8C93B37F9B89228 CRC64;

Query Match 65.8%; Score 89.5; DB 5; Length 970;
Best Local Similarity 66.7%; Pred. No. 0.00069;
Matches 14; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

QY 1 KHKHKHKKHKKH-KHKKHK 20
Db 46 RRRHHKKHKKHKKHKKHKK 66

```

Search completed: January 20, 2004, 18:26:24
 Job time : 27.6377 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 18:05:34 ; Search time 31.0145 Seconds
(without alignments)
102.356 Million cell updates/sec

Title: US-10-018-103A-6

Perfect score: 136

Sequence: 1 KHHHKKHHKKHHKKHHKK 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A Geneseq_19Jun03.*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	136	100.0	20	AAE06231	Histidine copolymer
2	97	71.3	49	22	Human liver peptid
3	97	71.3	49	22	Peptide #4697 enco
4	97	71.3	49	22	Peptide #4801 enco
5	97	71.3	49	22	Protein #4583 enco
6	97	71.3	49	22	Human brain expres
7	97	71.3	49	22	Human bone marrow
8	97	71.3	49	22	Peptide #4697 enco
9	97	71.3	49	22	Peptide #4792 enco

10	97	71.3	49	22	AAW05877	Peptide #4559 enco
11	97	71.3	49	23	ABG40067	Human peptid enco
12	97	71.3	52	22	ABG17093	Novel human diagno
13	97	71.3	292	22	ABG52171	Human liver peptid
14	97	71.3	292	22	ABG32108	Peptide #4759 enco
15	97	71.3	292	22	ABG37360	Peptide #4866 enco
16	97	71.3	292	22	ABG22646	Protein #4645 enco
17	97	71.3	292	22	AAW58050	Human brain expres
18	97	71.3	292	22	AAW70490	Human bone marrow
19	97	71.3	292	22	AAW18326	Peptide #4760 enco
20	97	71.3	292	22	AAW30814	Peptide #4851 enco
21	97	71.3	292	22	AAW05933	Peptide #4615 enco
22	97	71.3	292	23	ABG40129	Human peptid enco
23	96	70.6	104	21	AAW24812	Plant SDF encoded
24	96	70.6	104	21	AAW44772	Arabidopsis thalia
25	96	70.6	195	21	AAW24811	Plant SDF encoded
26	96	70.6	195	21	AAW44771	Arabidopsis thalia
27	96	70.6	233	22	ABG19767	Novel human diagno
28	96	70.6	407	23	ABW92012	Herbicidally activ
29	95	69.9	95	22	ABG03475	Novel human diagno
30	94	69.1	134	22	ABG05192	Novel human diagno
31	94	69.1	156	22	ABG05190	Novel human diagno
32	94	69.1	183	22	ABG25337	Novel human diagno
33	94	69.1	231	22	ABG05191	Novel human diagno
34	93	68.4	96	20	AAW99830	HIV vpr protein se
35	93	68.4	96	20	AAW99834	HIV A59p protein s
36	93	68.4	96	20	AAW99835	HIV L68S protein s
37	93	68.4	96	20	AAW99836	HIV H71C protein s
38	93	68.4	96	20	AAW99837	HIV G75A protein s
39	93	68.4	96	20	AAW99838	HIV C76S protein s
40	92	67.6	39	21	AAW12087	Arabidopsis thalia
41	92	67.6	45	21	AAW12086	Arabidopsis thalia
42	92	67.6	59	22	ABG05885	Novel human diagno
43	92	67.6	82	22	ABG47362	Human liver peptid
44	92	67.6	82	22	ABG27363	Human peptid #14
45	92	67.6	82	22	ABG32509	Peptide #15 encode

ALIGNMENTS

RESULT 1
AAE06231
ID AAE06231 standard; peptide; 20 AA.

AC AAE06231;

XX 25-SEP-2001 (first entry)

XX Histidine copolymer, HHH-K peptide.

XX Histidine copolymer; enzyme deficiency; phenylalanine hydroxylase;
XX ornithine transcarbamylase; adenosine deaminase; tyrosine hydroxylase;
XX purine nucleoside phosphorylase; alpha1-antitrypsin; apolipoprotein E;
XX chronic granulomatous disease; sickle cell; beta-thalassemia; anaemia;
XX familial hypercholesterolaemia; low-density lipoprotein receptor;
XX phenylketonuria; Faconi's anaemia; haemophilia; muscular dystrophy;
XX cystic fibrosis; Parkinson's disease; retinitis pigmentosa; cytostatic;
XX lysosomal storage disease; mycoplasma; type 1 disease; cardiant;
XX diabetic retinopathy; human immunodeficiency virus disease; infection;
XX cardiac disease; peripheral vascular disease; arthritis;
XX nontropic; haemostatic; virucide; gene therapy.

XX Unidentified.

OS WO200147496-A1.

XX 05-JUL-2001.

XX 20-DEC-2000; 2000WO-US34603.

XX 29-DEC-1999; 99US-0173576.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-496933/54.
 XX New spatially-addressable set of single exon nucleic acid probes,
 PT useful for measuring gene expression in sample derived from human
 PT breast, comprises number of single exon nucleic acid probes -
 XX
 XX Claim 27; SEQ ID NO 15014; 327pp + sequence listing; English.
 XX
 XX The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and Bt 474 cells. The method involves contacting
 CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for
 CC determining predisposition and/or prognosing breast disease. Gene
 CC expression analysis is useful for assessing the toxicity of chemical
 CC agents on cells. The microarray of this invention presents a far greater
 CC diversity of probes for measuring gene expression, with far less bias
 CC than expressed sequence tag microarrays. The method is suitable for
 CC rapid production of functional information from genomic sequence. The
 CC present sequence is a peptide encoded by a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 49 AA;
 Query Match 71.3%; Score 97; DB 22; Length 49;
 Best Local Similarity 70.6%; Pred. No. 6.7e-06;
 Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 3 HHKHHKHHKHHKHHK 19
 Db 13 HHHHHHHHHHHHHHR 29
 RESULT 4
 ABB37295
 ID ABB37295 standard; Peptide; 49 AA.
 AC ABB37295;
 XX
 XX 04-FEB-2002 (first entry)
 DE Peptide #4801 encoded by human foetal liver single exon probe.
 XX Human; foetal liver; gene expression; single exon nucleic acid probe.
 KW Homo sapiens.
 OS WO200157277-A2.
 PN 09-AUG-2001.
 PD 30-JAN-2001; 2001WO-US00669.
 PF 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488899/53.
 XX Single exon nucleic acid probes for analyzing gene expression in human
 PT hearts -
 PT
 XX Claim 15; SEQ ID No 24354; 530pp; English.
 XX The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart (see
 CC ABA21535-ABA41305). The present sequence is a protein encoded by one such
 CC probe. The probes may be used for predicting, measuring and displaying

DR WPI; 2001-483447/52.
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver -
 XX
 XX Claim 27; SEQ ID NO 29930; 639pp + sequence listing; English.
 XX
 XX The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC fetal liver. The present sequence is a peptide encoded by a single exon
 CC nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 49 AA;
 Query Match 71.3%; Score 97; DB 22; Length 49;
 Best Local Similarity 70.6%; Pred. No. 6.7e-06;
 Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 3 HHKHHKHHKHHKHHK 19
 Db 13 HHHHHHHHHHHHHHR 29
 RESULT 5
 ABB22584
 ID ABB22584 standard; Protein; 49 AA.
 AC ABB22584;
 XX
 XX 23-JAN-2002 (first entry)
 DE Protein #4583 encoded by probe for measuring heart cell gene expression.
 XX Human; gene expression; heart; microarray; vascular system;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease.
 XX Homo sapiens.
 OS WO200157274-A2.
 PN 09-AUG-2001.
 PD 30-JAN-2001; 2001WO-US00666.
 PF 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488899/53.
 XX Single exon nucleic acid probes for analyzing gene expression in human
 PT hearts -
 PT
 XX Claim 15; SEQ ID No 24354; 530pp; English.
 XX The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart (see
 CC ABA21535-ABA41305). The present sequence is a protein encoded by one such
 CC probe. The probes may be used for predicting, measuring and displaying

CC gene expression in samples derived from the human heart via microarrays.
CC By measuring gene expression, the probes are useful for predicting, the
CC diagnosing, grading, staging, monitoring and prognosing diseases of the
CC human heart and vascular system e.g. cardiovascular disease,
CC hypertension, cardiac arrhythmias and congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 49 AA;

Query Match 71.3%; Score 97; DB 22; Length 49;
Best Local Similarity 70.6%; Pred. No. 6.7e-06;
Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 3 HHHKHHKHHKHHKHHK 19
Dy 13 HHHKHHKHHKHHKHHK 29

RESULT 6
AAM57997
ID AAM57997 standard; Protein; 49 AA.

XX AC AAM57997;
XX DT 05-NOV-2001 (first entry)

XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 30102.
XX DE Human; brain expressed exon; gene expression analysis; probe;
XX KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX KW epilepsy; cancer.

XX OS Homo sapiens.
XX FN WO200157275-A2.
XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00667.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.

XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX brains -

XX PS Example 4; SEQ ID NO: 30102; 650pp + Sequence Listing; English.

XX CC The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, f
XX epilepsy and cancers. The present sequence is a protein encoded by one of
XX the probes of the invention.

XX SQ Sequence 49 AA;

Query Match 71.3%; Score 97; DB 22; Length 49;
Best Local Similarity 70.6%; Pred. No. 6.7e-06;

Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
Qy 3 HHHKHHKHHKHHKHHK 19
Dy 13 HHHKHHKHHKHHKHHK 29

RESULT 7
AAM70433
ID AAM70433 standard; Protein; 49 AA.

XX AC AAM70433;
XX DT 06-NOV-2001 (first entry)

XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 30739.
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma.

XX OS Homo sapiens.

XX FN WO200157276-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00668.

XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488900/53.

XX PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -

XX PS Example 4; SEQ ID NO: 30739; 658pp + Sequence Listing; English.

XX CC The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention.

XX SQ Sequence 49 AA;

Query Match 71.3%; Score 97; DB 22; Length 49;
Best Local Similarity 70.6%; Pred. No. 6.7e-06;
Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 3 HHHKHHKHHKHHKHHK 19
Dy 13 HHHKHHKHHKHHKHHK 29

RESULT 8
AAM18263
ID AAM18263 standard; Protein; 49 AA.

XX AC AAM18263;

XX DT 12-OCT-2001 (first entry)

DE Peptide #4697 encoded by probe for measuring cervical gene expression.
 XX Probe; human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer.
 XX

OS Homo sapiens.

XX WO200157278-A2.

PN 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00670.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human cervical epithelial cells -

XX Claim 27; SEQ ID No 23089; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes
 CC (SENP; see AAI10068-AA128459). The present sequence is a peptide encoded
 CC by one such probe. The SENPs are derived from human Hela cells. The SENPs
 CC can be used to produce a single exon microarray, which can be used for
 CC measuring human gene expression in a sample derived from human cervical
 CC epithelial cells. By measuring gene expression, the probes are therefore
 CC useful in grading and/or staging of diseases of the cervix, notably
 CC cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 49 AA;

Query Match 71.3%; Score 97; DB 22; Length 49;
 Best Local Similarity 70.6%; Pred. No. 6.7e-06;
 Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 HHKHHKHHKHHKHHK 19

Db 13 HHHHHHHHHHHHHH 29

RESULT 9

AAM30755

ID AAM30755 standard; Protein; 49 AA.

XX AAM30755;

XX 17-OCT-2001 (first entry)

XX Peptide #4792 encoded by probe for measuring placental gene expression.

XX Probe; microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder.

XX Homo sapiens.

XX WO200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00663.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human placenta -

XX Claim 27; SEQ ID No 31024; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENP;

CC see AAI31315-AA157546). The present sequence is a peptide encoded by one

CC such probe. The probes are useful for producing a microarray for

CC predicting, measuring and displaying gene expression in samples derived

CC from human placenta. The probes are useful for antenatal diagnosis of

CC human genetic disorders.

XX Sequence 49 AA;

Query Match 71.3%; Score 97; DB 22; Length 49;

Best Local Similarity 70.6%; Pred. No. 6.7e-06;

Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 HHKHHKHHKHHKHHK 19

Db 13 HHHHHHHHHHHHHH 29

RESULT 10

AAM05877

ID AAM05877 standard; Protein; 49 AA.

XX AAM05877;

XX 09-OCT-2001 (first entry)

XX Peptide #4559 encoded by probe for measuring breast gene expression.
 DE Probe; human; breast disease; breast cancer; development disorder;
 XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX Homo sapiens.

XX WO200157270-A2.

XX 09-AUG-2001.

XX 29-JAN-2001; 2001WO-US00661.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX 09-AUG-2001.

us-10-018-103a-6.rag

Wed Jan 21 11:27:53 2004

PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX Claim 27; SEQ ID NO 29995; 639pp + sequence listing; English.
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 292 AA;
Query Match 71.3%; Score 97; DB 22; Length 292;
Best Local Similarity 66.7%; Pred. No. 4e-05;
Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 3 HHHKHHKHHKHHKHHK 20
DB 81 HHHHHHHHHHHHHHHRR 98
Search completed: January 20, 2004, 18:23:19
Job time : 32.0145 secs

PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-496933/54.
XX New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX Claim 27; SEQ ID NO 15076; 327pp + sequence listing; English.
XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a peptide encoded by a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 292 AA;
Query Match 71.3%; Score 97; DB 22; Length 292;
Best Local Similarity 66.7%; Pred. No. 4e-05;
Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 3 HHHKHHKHHKHHKHHK 20
DB 81 HHHHHHHHHHHHHHHRR 98
RESULT 15
ABB37360
ID ABB37360 standard; Peptide; 292 AA.
XX AC ABB37360;
XX DT 04-FEB-2002 (first entry)
XX DE Peptide #4866 encoded by human foetal liver single exon probe.
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX OS Homo sapiens.
XX PN WO200157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00669.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 18:22:04 ; Search time 11.3043 Seconds
(without alignments)
74.858 Million cell updates/sec

Title: US-10-018-103A-6
Perfect score: 136
Sequence: 1 KKHKKHHKKHHKKHHKK 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- Issued Patents AA:*
- 1: /cgn2_6/ptodata/1/iaa/5A COMB.pep.*
 - 2: /cgn2_6/ptodata/1/iaa/5B COMB.pep.*
 - 3: /cgn2_6/ptodata/1/iaa/6A COMB.pep.*
 - 4: /cgn2_6/ptodata/1/iaa/6B COMB.pep.*
 - 5: /cgn2_6/ptodata/1/iaa/PTUS COMB.pep.*
 - 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	93	68.4	363	4	US-09-328-352-4930
2	86	63.2	303	2	US-08-203-532F-2
3	86	63.2	303	3	US-09-078-465-2
4	86	63.2	303	5	PCT-US95-01882A-2
5	84	61.8	1199	3	US-09-208-742-2
6	84	61.8	1199	4	US-09-332-295-4
7	84	61.8	1199	4	US-09-709-979-4
8	83	61.0	368	1	US-08-211-942-17
9	82	60.3	60	1	US-08-255-457-1
10	82	60.3	60	2	US-09-115-032-1
11	82	60.3	60	5	PCT-US95-05772-1
12	81.5	59.9	763	2	US-08-677-862-2
13	81.5	59.9	763	2	US-09-252-571-2
14	81.5	59.9	763	3	US-09-434-065-2
15	81.5	59.9	763	3	US-08-789-275-4
16	81.5	59.9	763	3	US-08-789-275-5
17	79	58.1	302	2	US-08-203-532F-4
18	79	58.1	302	3	US-08-950-860-16
19	79	58.1	302	3	US-09-078-465-4
20	79	58.1	302	5	PCT-US95-01882A-4
21	75.5	55.5	434	4	US-09-252-991A-30855
22	75	55.1	480	1	US-07-882-292-2
23	75	55.1	480	2	US-08-331-644-2
24	75	55.1	480	5	PCT-US93-04102-2
25	75	55.1	633	3	US-08-557-006C-43
26	74	54.4	1664	1	US-09-599-652-2
27	74	54.4	1664	2	US-08-642-846-2

28	74	54.4	1664	4	US-09-264-604-2	Sequence 2, Appli
29	73.5	54.0	313	3	US-08-686-528A-3	Sequence 3, Appli
30	73.5	54.0	313	3	US-09-456-287-3	Sequence 3, Appli
31	73.5	54.0	337	3	US-08-686-528A-2	Sequence 2, Appli
32	73.5	54.0	337	3	US-09-456-287-2	Sequence 2, Appli
33	73	53.7	349	3	US-09-461-474-12	Sequence 12, Appli
34	72	52.9	414	5	PCT-US92-06840-2	Sequence 2, Appli
35	72	52.9	431	1	US-08-311-023-2	Sequence 2, Appli
36	71	52.2	448	3	US-09-461-474-8	Sequence 8, Appli
37	70	51.5	765	4	US-09-252-991A-17416	Sequence 17416, A
38	70	51.5	1182	3	US-09-041-886-21	Sequence 21, Appli
39	70	51.5	2265	2	US-08-149-097D-36	Sequence 36, Appli
40	70	51.5	2509	2	US-08-149-097D-35	Sequence 35, Appli
41	69.5	51.1	115	4	US-09-461-325-160	Sequence 160, Appli
42	68.5	50.4	115	4	US-09-996-243-95	Sequence 95, Appli
43	68.5	50.4	726	3	US-09-126-980-2	Sequence 2, Appli
44	68.5	50.4	726	3	US-09-476-482-2	Sequence 2, Appli
45	68.5	50.4	726	4	US-09-517-603-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-328-352-4930
; Sequence 4930, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4930
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4930

Query Match 68.4%; Score 93; DB 4; Length 363;
Best Local Similarity 70.6%; Pred. NO. 0.00013;
Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 3 HHHKKHHKKHHKKHHK 19
Db 188 HHHHHHHHHHHHHHHE 204

RESULT 2
US-08-203-532F-2
; Sequence 2, Application US/08203532F
; Patent No. 5856121
; GENERAL INFORMATION:
; APPLICANT: Gorski, David H.
; APPLICANT: Walsh, Kenneth
; TITLE OF INVENTION: Growth Arrest Homeobox Gene
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calfee, Halter, and Griswold
; STREET: 800 Superior Avenue
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: U.S.A.
; ZIP: 44114-2688
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/203,532F

Wed Jan 21 11:27:53 2004

```

; FILING DATE: 435
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Golrick, Mary E.
; REGISTRATION NUMBER: 34829
; REFERENCE/DOCKET NUMBER: 22311/00114
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (216) 622-8200
; TELEFAX: (216) 241-0816
; TELEX: 980499
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 303 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-203-532F-2

Query Match 63.2%; Score 86; DB 2; Length 303;
Best Local Similarity 57.9%; Pred. No. 0.00075;
Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 KHHKHHKHHKHHKHHK 20
Db 63 QHRRGHHHHHHHHHHQ 81

RESULT 3
US-09-078-465-2
; Sequence 2, Application US/09078465
; Patent No. 6280969
; GENERAL INFORMATION:
; APPLICANT: Gorski, David H.
; APPLICANT: Walsh, Kenneth
; TITLE OF INVENTION: Growth Arrest Homeobox Gene
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calfee, Halter, and Griswold
; STREET: 800 Superior Avenue
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: U.S.A.
; ZIP: 44114-2688
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Golrick, Mary E.
; REGISTRATION NUMBER: 34829
; REFERENCE/DOCKET NUMBER: 22311/00114
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (216) 622-8200
; TELEFAX: (216) 241-0816
; TELEX: 980499
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 303 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-01882A-2

Query Match 63.2%; Score 86; DB 5; Length 303;
Best Local Similarity 57.9%; Pred. No. 0.00075;
Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 KHHKHHKHHKHHKHHK 20
Db 63 QHRRGHHHHHHHHHHQ 81

RESULT 4
PCT-US95-01882A-2
; Sequence 2, Application PC/TUS9501882A
; GENERAL INFORMATION:
; APPLICANT: Gorski, David H.
; APPLICANT: Walsh, Kenneth
; TITLE OF INVENTION: Growth Arrest Homeobox Gene
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calfee, Halter, and Griswold
; STREET: 800 Superior Avenue
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: U.S.A.
; ZIP: 44114-2688
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Golrick, Mary E.
; REGISTRATION NUMBER: 34829
; REFERENCE/DOCKET NUMBER: 22311/00114
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (216) 622-8200
; TELEFAX: (216) 241-0816
; TELEX: 980499
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 303 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-01882A-2

Query Match 63.2%; Score 86; DB 5; Length 303;
Best Local Similarity 57.9%; Pred. No. 0.00075;
Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 KHHKHHKHHKHHKHHK 20
Db 63 QHRRGHHHHHHHHHHQ 81

RESULT 5
US-09-208-742-2
; Sequence 2, Application US/09208742
; Patent No. 6174679
; GENERAL INFORMATION:
; APPLICANT: Kaufmann, Joerg
; TITLE OF INVENTION: Cif150/HTAFII150 is Necessary for Cell
; FILE REFERENCE: 1453.002
; CURRENT APPLICATION NUMBER: US/09/208,742
; CURRENT FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1199
; TYPE: PRT
; ORGANISM: human

```


US-09-208-742-2

Query Match 61.8%; Score 84; DB 3; Length 1199;
Best Local Similarity 60.9%; Pred. NO. 0.0049;
Matches 14; Conservative 0; Mismatches 3; Indels

Qy 3 HHHKNNH-----KKNNHNHNK 19
 ||| ||| ||| ||| |||
Db 1146 HHHHHNHNHNKKKKKKNNHNHNK 1168

RESULT 6

US-09-332-233-4
: Sequence 4. Application US/09332295

```

; PATENT NO: 0505372
; GENERAL INFORMATION:
; APPLICANT: Kaufmann, Joerg
; TITLE OF INVENTION: CF130 INHIBITS CELL CYCLE PROGRESSION
; FILE REFERENCE: 200130.456 / 1513.003
; CURRENT APPLICATION NUMBER: US/09/332,295
; CURRENT FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1199
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-332-295-4

```

Query Match 61.8%; Score 84; DB 4; Length 1199;
Best Local Similarity 60.9%; Pred. No. 0.0049;
Matches 14; Conservative 0; Mismatches 3; Indels

Qy 3 HNNKNN-----KGNHNNHNNK 19
1146 HNNNNNNHNNKKKKKKHNNKNNK 1168
Db

RESULT 7

US-09-709-373-4 : Sequence 4 Application IIS/09709979

```

; FACULT NO. 8423822
;
; GENERAL INFORMATION:
;
; APPLICANT: Kaufmann, Joerg
;
; TITLE OF INVENTION: CFF130 INHIBITS CELL CYCLE PROGRESSION
;
; FILE REFERENCE: 200130.456 / 1513.003
;
; CURRENT APPLICATION NUMBER: US/09/709,979
;
; CURRENT FILING DATE: 2000-11-09
;
; PRIOR APPLICATION NUMBER: US 09/332,295
;
; PRIOR FILING DATE: 1999-06-11
;
; NUMBER OF SEQ ID NOS: 4
;
; SOFTWARE: FastSEQ for Windows Version 3.0
;
; SEQ ID NO 4
;
; LENGTH: 1199
;
; TYPE: PRT
;
; ORGANISM: Homo sapien
;
; US-09-709-979-4

```

Query Match 61.8%; Score 84; DB 4; Length 1199;
Best Local Similarity 60.9%; Pred. No. 0.0049;
Matches 14: Conservative 0; Mismatches 3; Indels

Qy 3 HHHKHHH-----KCHHHKHHHK 19
1146 HHHHHHHHHKKKKCHKHKKHK 1168
Db

RESULTS

US-08-211-942-1 / : Sequence 17 Application 118/08211942

; Patent No. 5523287
 ; GENERAL INFORMATION:
 ; APPLICANT: Friedrich. Thomas

APPLICANT: Bialojan, Siegfried
 APPLICANT: Kroeger, Burkhard
 APPLICANT: Kuenast, Christoph
 TITLE OF INVENTION: No. 5523287e1 thrombin-inhibitory protein from asessasin
 TITLE OF INVENTION: bugs.
 NUMBER OF SEQUENCES: 22
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Keil & Weinkauf
 STREET: 1101 Connecticut Avenue
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
 COMPUTER: IBM AT-compatible, 80486 processor
 OPERATING SYSTEM: MS-DOS version 6.0
 SOFTWARE: Wordperfect version 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/211,942
 FILING DATE:
 CLASSIFICATION: 435
 CLASSIFICATION: C07K 13/00
 CLASSIFICATION: A61K 37/64
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/EP92/02450
 FILING DATE: 27-OCT-1992
 INFORMATION FOR SEQ ID NO: 17:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 368 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 US-08-211-942-17

Query Match 61.0%; Score 83; DB 1; Length 368;
Best Local Similarity 64.7%; Pred. No. 0.0021;
Matches 11; Conservative 1; Mismatches 5; Indels

Qy	2	K	H	H	H	K	K	H	H	K	H	H	18
Db	176	Q	H	H	H	H	H	P	P	P	H	H	19

RESULT 9

US-08-255-457-1
 ; Sequence 1, Application US/08255457
 ; Patent No. 5780040
 ; GENERAL INFORMATION:
 ; APPLICANT: Plaut, Andrew G.
 ; APPLICANT: Gilbert-Rothstein, Joanne V.
 ; APPLICANT: Wright, Andrew
 ; TITLE OF INVENTION: HELICOBACTER PYLORI NICKEL BINDING
 ; TITLE OF INVENTION: PROTEIN
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSER: Fish & Richardson
 ; STREET: 225 Franklin Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02110-2804
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/255,457
 ; FILING DATE:
 ; CLASSIFICATION: 530
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Clark, Paul C.
 ; REGISTRATION NUMBER: 30,162

REFERENCE/DOCKET NUMBER: 00398/090001
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 60 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-255-457-1

Query Match 60.3%; Score 82; DB 1; Length 60;
Best Local Similarity 60.0%; Pred. No. 0.00048;
Matches 12; Conservative 1; Mismatches 3; Indels 1;
Gaps 4;

Qy 3 HHHKHHKHHH----KHHH 18
Db 11 HHHHHHTHHHHVHGGEHHH 30

RESULT 10
US-09-115-032-1
Sequence 1, Application US/09115032
Patent No. 5972348
GENERAL INFORMATION:
APPLICANT: Plaut, Andrew G.
APPLICANT: Gilbert-Rothstein, Joanne V.
TITLE OF INVENTION: HELICOBACTER PYLORI NICKEL BINDING
TITLE OF INVENTION: PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/115,032
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/255,457
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul C.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00398/090001
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 60 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-115-032-1

Query Match 60.3%; Score 82; DB 2; Length 60;
Best Local Similarity 60.0%; Pred. No. 0.00048;
Matches 12; Conservative 1; Mismatches 3; Indels 1;
Gaps 4;

Qy 3 HHHKHHKHHH----KHHH 18
Db 11 HHHHHHTHHHHVHGGEHHH 30

RESULT 11
PCT-US95-05772-1
Sequence 1, Application PC/TUS9505772
GENERAL INFORMATION:
APPLICANT: Plaut, Andrew G.
APPLICANT: Gilbert-Rothstein, Joanne V.
TITLE OF INVENTION: HELICOBACTER PYLORI NICKEL
TITLE OF INVENTION: BINDING PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05772
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul C.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00398/090001
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 60 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-05772-1

Query Match 60.3%; Score 82; DB 5; Length 60;
Best Local Similarity 60.0%; Pred. No. 0.00048;
Matches 12; Conservative 1; Mismatches 3; Indels 1;
Gaps 4;

Qy 3 HHHKHHKHHH----KHHH 18
Db 11 HHHHHHTHHHHVHGGEHHH 30

RESULT 12
US-08-677-862-2
Sequence 2, Application US/08677862
Patent No. 5874230
GENERAL INFORMATION:
APPLICANT: SONG, Ho Yeong
APPLICANT: ROTH, Mike
TITLE OF INVENTION: TRAF2-Associated Kinase
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, 34th floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/677,862
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brezner, David J
; REGISTRATION NUMBER: 24,774
; REFERENCE/DOCKET NUMBER: T96-005/A63613
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415 781-1989
; TELEFAX: 415 398-3249
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-677-862-2

Query Match          59.9%; Score 81.5; DB 2; Length 763;
Best Local Similarity 63.2%; Pred. No. 0.0063;
Matches 12; Conservative 0; Mismatches 4; Indels 3; Gaps 1;

QY 3 HHH---KHHHKKHHKHHH 18
   ||| ||| ||| ||| |||
Db 600 HHGNGSSHHHHHHHHHH 618

RESULT 13
US-09-252-571-2
; Sequence 2, Application US/09252571
; Patent No. 5981250
; GENERAL INFORMATION:
; APPLICANT: SONG, Ho Yeong
; APPLICANT: ROTHE, Mike
; TITLE OF INVENTION: TRAP2-Associated Kinase
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, 34th floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/252,571
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/677,862
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brezner, David J
; REGISTRATION NUMBER: 24,774
; REFERENCE/DOCKET NUMBER: T96-005/A63613
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415 781-1989
; TELEFAX: 415 398-3249
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
```

```
;
; MOLECULE TYPE: peptide
; US-09-252-571-2

Query Match          59.9%; Score 81.5; DB 2; Length 763;
Best Local Similarity 63.2%; Pred. No. 0.0063;
Matches 12; Conservative 0; Mismatches 4; Indels 3; Gaps 1;

QY 3 HHH---KHHHKKHHKHHH 18
   ||| ||| ||| ||| |||
Db 600 HHGNGSSHHHHHHHHHH 618

RESULT 14
US-09-434-065-2
; Sequence 2, Application US/09434065
; Patent No. 6107074
; GENERAL INFORMATION:
; APPLICANT: SONG, Ho Yeong
; APPLICANT: ROTHE, Mike
; TITLE OF INVENTION: TRAP2-Associated Kinase
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, 34th floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/434,065
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/677,862
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brezner, David J
; REGISTRATION NUMBER: 24,774
; REFERENCE/DOCKET NUMBER: T96-005/A63613
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415 781-1989
; TELEFAX: 415 398-3249
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-434-065-2

Query Match          59.9%; Score 81.5; DB 3; Length 763;
Best Local Similarity 63.2%; Pred. No. 0.0063;
Matches 12; Conservative 0; Mismatches 4; Indels 3; Gaps 1;

QY 3 HHH---KHHHKKHHKHHH 18
   ||| ||| ||| ||| |||
Db 600 HHGNGSSHHHHHHHHHH 618

RESULT 15
US-08-789-275-4
; Sequence 4, Application US/08789275A
; Patent No. 6251664
; GENERAL INFORMATION:
; APPLICANT: Palreja, Estivill Xavier
; APPLICANT: Pritchard, Melanie
; APPLICANT: Vilaro, Jordi Guimera
```

Wed Jan 21 11:27:53 2004

```
; TITLE OF INVENTION: Human Gene Sequence of the Down Syndrome Critical
; Patent No. 6251664
; TITLE OF INVENTION: Region of Human Chromosome 21, Coding for A Serine-
; TITLE OF INVENTION: Threonine Protein Kinase (MNB), Expressed in the
; TITLE OF INVENTION: Neuronal Regions Affected in Down Syndrome
; FILE REFERENCE: U 01114-4
; CURRENT APPLICATION NUMBER: US/08/789,275A
; CURRENT FILING DATE: 1997-01-28
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 4
; LENGTH: 763
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
;
US-08-789-275-4
Query Match      59.8; Score 81.5; DB 3; Length 763;
Best Local Similarity 63.2; Pred. No. 0.0063;
Matches 12; Conservative 0; Mismatches 4; Indels 3; Gaps 1;

Qy      3 HHH---KHHKHHKHHKHH 18
      ||| ||| ||| ||| |||
Db      600 HHGCGSSHHHHHHHHHHH 618

Search completed: January 20, 2004, 18:27:12
Job time : 11.3043 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 18:23:25 ; Search time 22.6087 Seconds
(without alignments)
180.887 Million cell updates/sec

Title: US-10-018-103A-6

Perfect score: 136

Sequence: 1 KKHKKHHKHHKHHKHHK 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 762491 seqs, 20448190 residues

Total number of hits satisfying chosen parameters: 762491

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	136	100.0	20	12	US-10-131-909A-6
2	136	100.0	20	15	US-10-018-103A-6
3	110.5	81.2	17	12	US-10-131-909A-17
4	98	72.1	378	12	US-10-029-386-33892
5	97	71.3	49	9	US-09-864-761-37882
6	97	71.3	292	9	US-09-864-761-37944
7	96	70.6	82	12	US-10-029-386-30014
8	92	67.6	29	9	US-09-864-761-33313
9	92	67.6	87	9	US-09-864-761-33727
10	92	67.6	87	9	US-09-864-761-34744
11	92	67.6	90	12	US-10-315-515-39
12	92	67.6	90	12	US-10-315-515-44
13	92	67.6	93	12	US-10-315-515-46
14	92	67.6	95	12	US-10-315-515-35
15	92	67.6	96	12	US-10-315-515-34

16	92	67.6	96	12	US-10-315-515-36	Sequence 36, Appl
17	92	67.6	96	12	US-10-315-515-37	Sequence 37, Appl
18	92	67.6	96	12	US-10-315-515-40	Sequence 40, Appl
19	92	67.6	96	12	US-10-315-515-41	Sequence 41, Appl
20	92	67.6	96	12	US-10-315-515-42	Sequence 42, Appl
21	92	67.6	96	12	US-10-315-515-45	Sequence 45, Appl
22	92	67.6	105	12	US-10-315-515-38	Sequence 43, Appl
23	92	67.6	106	12	US-10-315-515-43	Sequence 38, Appl
24	92	67.6	523	12	US-10-017-161-1982	Sequence 1982, Ap
25	92	67.6	523	12	US-10-292-798-1630	Sequence 1630, Ap
26	89.5	65.8	21	12	US-10-131-909A-9	Sequence 9, Appl
27	89.5	65.8	21	12	US-10-131-909A-16	Sequence 16, Appl
28	89.5	65.8	21	15	US-10-018-103A-9	Sequence 9, Appl
29	89.5	65.8	21	15	US-10-018-103A-16	Sequence 16, Appl
30	89	65.4	461	10	US-09-764-868-765	Sequence 765, App
31	87	64.0	37	9	US-09-864-761-40909	Sequence 40909, A
32	86.5	63.6	16	12	US-10-104-307-17	Sequence 17, Appl
33	86.5	63.6	485	12	US-10-295-027-476	Sequence 476, App
34	86	63.2	186	12	US-10-029-386-34005	Sequence 34005, A
35	86	63.2	303	12	US-09-940-673-2	Sequence 2, Appl
36	85	62.5	30	9	US-09-864-761-36251	Sequence 36251, A
37	85	62.5	78	9	US-09-864-761-37352	Sequence 37352, A
38	85	62.5	108	12	US-10-029-386-31185	Sequence 31185, A
39	85	62.5	156	9	US-09-864-761-41679	Sequence 41679, A
40	85	62.5	625	9	US-09-853-386-63	Sequence 63, Appl
41	85	62.5	625	12	US-10-414-080-13	Sequence 13, Appl
42	85	62.5	626	9	US-09-853-386-64	Sequence 64, Appl
43	85	62.5	626	9	US-09-853-386-65	Sequence 65, Appl
44	85	62.5	626	9	US-09-853-386-96	Sequence 96, Appl
45	85	62.5	626	12	US-10-414-080-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1

US-10-131-909A-6
; Sequence 6, Application US/10131909A
; Publication No. US20030165567A1
; GENERAL INFORMATION:
; APPLICANT: Mixson, A. James
; TITLE OF INVENTION: Histidine Copolymer and Methods For Using Same
; FILE REFERENCE: 5627*6
; CURRENT APPLICATION NUMBER: US/10/131,909A
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US 10/018103
; PRIOR FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: PCT/US00/34603
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/173576
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
US-10-131-909A-6

Query Match 100.0%; Score 136; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.6e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKHKKHHKHHKHHKHHK 20

|||||

Db 1 KKHKKHHKHHKHHKHHK 20

|||||

RESULT 2

US-10-018-103A-6
; Sequence 6, Application US/10018103A

Wed Jan 21 11:27:53 2004

```
; Publication No. US20030045465A1
; GENERAL INFORMATION:
; APPLICANT: Mixson, A. James
; TITLE OF INVENTION: Histidine Copolymer and Methods For Using Same
; FILE REFERENCE: 5627*5
; CURRENT APPLICATION NUMBER: US/10/018,103A
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US 60/173576
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: PCT/US00/34603
; PRIOR FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
US-10-018-103A-6

Query Match      100.0%; Score 136; DB 15; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.6e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KHHHHKHHKHHKHHKHHK 20
Db 1 KHHHHKHHKHHKHHKHHK 20

RESULT 3
US-10-131-909A-17
; Sequence 17, Application US/10131909A
; Publication No. US20030165567A1
; GENERAL INFORMATION:
; APPLICANT: Mixson, A. James
; TITLE OF INVENTION: Histidine Copolymer and Methods For Using Same
; FILE REFERENCE: 5627*6
; CURRENT APPLICATION NUMBER: US/10/131,909A
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US 10/018103
; PRIOR FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: PCT/US00/34603
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/173576
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17
; LENGTH: 17
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
US-10-131-909A-17

Query Match      81.2%; Score 110.5; DB 12; Length 17;
Best Local Similarity 94.4%; Pred. No. 1.2e-05;
Matches 17; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 2 KHHHHKHHKHHKHHKHHK 19
Db 1 KHHHKHHH-KHHHKHHK 17

RESULT 4
US-10-029-386-33892
; Sequence 33892, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
```

```
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: ACOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33892
; LENGTH: 378
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP001751.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.6
US-10-029-386-33892

Query Match      72.1%; Score 98; DB 12; Length 378;
Best Local Similarity 75.0%; Pred. No. 0.0032;
Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 HHHKHHKHHKHHKHHK 18
Db 331 HHHQHHNNHHHHQHH 346

RESULT 5
US-09-864-761-37882
; Sequence 37882, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: ACOMICA-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
```

;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 37882
;; LENGTH: 49
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC011186.2
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.3
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
US-09-864-761-37882

Query Match 71.3%; Score 97; DB 9; Length 49;
Best Local Similarity 70.6%; Pred. No. 0.00071;
Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 HHHKHHKHHKHHKHHK 19
|||:|:| ||| |||:
DB 13 HHHRRHHHHHHHHRR 29

RESULT 6

US-09-864-761-37944
;; Sequence 37944, Application US/09864761
;; Patent No. US20020048763A1
;; GENERAL INFORMATION:
;; APPLICANT: Penn, Sharron G.
;; APPLICANT: Hanzel, David K.
;; APPLICANT: Chen, Wensheng
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;; FILE REFERENCE: Aemica-X-1
;; CURRENT APPLICATION NUMBER: US/09/864,761
;; CURRENT FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 37944
;; LENGTH: 292
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC005414.2
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.5
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 6.2
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.1
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 6.1
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.5
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.2
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 7.7
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 7.9
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 11
US-09-864-761-37944

Query Match 71.3%; Score 97; DB 9; Length 292;
Best Local Similarity 66.7%; Pred. No. 0.0032;
Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 HHHKHHKHHKHHKHHK 20
|||:|:| ||| |||:
DB 81 HHHRRHHHHHHHHRR 98

RESULT 7

US-10-029-386-30014
;; Sequence 30014, Application US/10029386
;; Publication No. US20030194704A1
;; GENERAL INFORMATION:
;; APPLICANT: Penn, Sharron G.
;; APPLICANT: Hanzel, David K.
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;; FILE REFERENCE: AEMICA-X-2
;; CURRENT APPLICATION NUMBER: US/10/029,386
;; CURRENT FILING DATE: 2001-12-20
;; NUMBER OF SEQ ID NOS: 34288
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 30014
;; LENGTH: 29
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL13895.4
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.88
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.8
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
US-10-029-386-30014

Query Match 70.6%; Score 96; DB 12; Length 29;
Best Local Similarity 66.7%; Pred. No. 0.00058;
Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Wed Jan 21 11:27:53 2004

QY 1 KKHKKHHKKHHKKHH 18
:|||||
Db 6 RRRHHHHHHHHHHHH 23

RESULT 8
US-09-864-761-33313
; Sequence 33313, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33313
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007263.3
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 8.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 7.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 8.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 9.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.2
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 7.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.1

OTHER INFORMATION: SWISSPROT HIT: Q12329, EVALUE 4.90e+00
US-09-864-761-33313

Query Match 67.6%; Score 92; DB 9; Length 82;
Best Local Similarity 75.0%; Pred. No. 0.0035;
Matches 12; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 HHHKKHHKKHHKKHH 18
:|||||
Db 59 HHHHHHHHHHHHHH 74

RESULT 9
US-09-864-761-33727
; Sequence 33727, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33727
; LENGTH: 87
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL096757.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1


```
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8
US-09-864-761-33727

Query Match          67.6%; Score 92; DB 9; Length 87;
Best Local Similarity 75.0%; Pred. No. 0.0037;
Matches 12; Conservative 0; Mismatches 4; Indels 0; Caps 0;

QY 3 HHHKHHKHHKHHKHH 18
   ||| ||| ||| |||
Db 62 HHHHHHHHHHHHHH 77

RESULT 10
US-09-864-761-34744
; Sequence 34744, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34744
; LENGTH: 87
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL080246.13
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.1
US-09-864-761-34744

Query Match          67.6%; Score 92; DB 9; Length 87;
Best Local Similarity 75.0%; Pred. No. 0.0037;
Matches 12; Conservative 0; Mismatches 4; Indels 0; Caps 0;

QY 3 HHHKHHKHHKHHKHH 18
   ||| ||| ||| |||
Db 62 HHHHHHHHHHHHHH 77

RESULT 11
US-10-315-515-39
; Sequence 39, Application US/10315515
; Publication No. US20030166190A1
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: NUCLEIC ACIDS RELATED TO PLANT
; TITLE OF INVENTION: RETROELEMENTS
; FILE REFERENCE: 08411-031001
; CURRENT APPLICATION NUMBER: US/10/315,515
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: US 60/339,060
; PRIOR FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 168
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Glycine max
US-10-315-515-39

Query Match          67.6%; Score 92; DB 12; Length 90;
Best Local Similarity 75.0%; Pred. No. 0.0038;
Matches 12; Conservative 0; Mismatches 4; Indels 0; Caps 0;

QY 3 HHHKHHKHHKHHKHH 18
   ||| ||| ||| |||
Db 37 HHHHHHHHHHHHHH 52

RESULT 12
US-10-315-515-44
; Sequence 44, Application US/10315515
; Publication No. US20030166190A1
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: NUCLEIC ACIDS RELATED TO PLANT
; TITLE OF INVENTION: RETROELEMENTS
; FILE REFERENCE: 08411-031001
; CURRENT APPLICATION NUMBER: US/10/315,515
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: US 60/339,060
; PRIOR FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 168
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 90
```

Wed Jan 21 11:27:53 2004

```
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-315-515-44
Query Match      67.6%; Score 92; DB 12; Length 90;
Best Local Similarity 75.0%; Pred. No. 0.0038;
Matches 12; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 HHHKHHKHHKHHH 18
   ||| ||| ||| |||
DB 37 HHHHHHHHHHHHHH 52

RESULT 13
US-10-315-515-46
; Sequence 46, Application US/10315515
; Publication No. US20030166190A1
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: NUCLEIC ACIDS RELATED TO PLANT
; TITLE OF INVENTION: RETROELEMENTS
; FILE REFERENCE: 08411-031001
; CURRENT APPLICATION NUMBER: US/10/315,515
; PRIOR FILING DATE: 2002-12-10
; PRIOR FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 168
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 46
; LENGTH: 93
; TYPE: PRT
; ORGANISM: Glycine max
US-10-315-515-46
Query Match      67.6%; Score 92; DB 12; Length 93;
Best Local Similarity 75.0%; Pred. No. 0.0039;
Matches 12; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 HHHKHHKHHKHHH 18
   ||| ||| ||| |||
DB 43 HHHHHHHHHHHHHH 58

RESULT 14
US-10-315-515-35
; Sequence 35, Application US/10315515
; Publication No. US20030166190A1
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: NUCLEIC ACIDS RELATED TO PLANT
; TITLE OF INVENTION: RETROELEMENTS
; FILE REFERENCE: 08411-031001
; CURRENT APPLICATION NUMBER: US/10/315,515
; PRIOR FILING DATE: 2002-12-10
; PRIOR FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 168
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-315-515-35
Query Match      67.6%; Score 92; DB 12; Length 95;
Best Local Similarity 75.0%; Pred. No. 0.004;
Matches 12; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 HHHKHHKHHKHHH 18
   ||| ||| ||| |||
DB 43 HHHHHHHHHHHHHH 58

RESULT 15
US-10-315-515-34
; Sequence 34, Application US/10315515
; Publication No. US20030166190A1
; GENERAL INFORMATION:
; APPLICANT: Wright, David A.
; APPLICANT: Voytas, Daniel F.
; TITLE OF INVENTION: NUCLEIC ACIDS RELATED TO PLANT
; TITLE OF INVENTION: RETROELEMENTS
; FILE REFERENCE: 08411-031001
; CURRENT APPLICATION NUMBER: US/10/315,515
; PRIOR FILING DATE: 2002-12-10
; PRIOR FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 168
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-315-515-34
Query Match      67.6%; Score 92; DB 12; Length 96;
Best Local Similarity 75.0%; Pred. No. 0.0041;
Matches 12; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 HHHKHHKHHKHHH 18
   ||| ||| ||| |||
DB 43 HHHHHHHHHHHHHH 58

Search completed: January 20, 2004, 18:28:43
Job time : 22.6087 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 18:18:04 ; Search time 10.1449 Seconds
(without alignments)
189.590 Million cell updates/sec

Title: US-10-018-103A-6
Perfect score: 136
Sequence: 1 KKHKKHHKKHHKKHHKK 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76:.*
1: p1r1:.*
2: p1r2:.*
3: p1r3:.*
4: p1r4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	96	70.6	407	G84783	probable pectinest
2	92	67.6	140	A54523	histidine-rich pro
3	92	67.6	351	1 KG20HL	histidine-rich gly
4	91	66.9	385	2 A84696	probable zinc tran
5	88	64.7	82	2 A29653	histidine-rich pro
6	87	64.0	1291	2 T13389	hypothetical prote
7	86	63.2	303	2 A56817	homeotic protein M
8	86	63.2	303	2 B49122	homeobox protein M
9	86	63.2	303	2 A48130	growth arrest-spec
10	85	62.5	430	2 S66671	neuron-derived rec
11	85	62.5	625	2 S71930	neuron-derived rec
12	85	62.5	628	2 JC2493	neuron derived orp
13	85	62.5	1212	2 T13804	shs protein - frui
14	84	61.8	326	2 D83483	probable metal tra
15	84	61.8	390	1 A38565	polycarb (pc) prot
16	84	61.8	580	2 T46024	hypothetical prote
17	83	61.0	420	2 T39712	hypothetical prote
18	83	61.0	427	2 T42516	hypothetical prote
19	83	61.0	1455	2 T23056	chromodomain helic
20	82.5	60.7	143	1 B64421	conserved hypothet
21	82.5	60.7	147	2 T16440	hypothetical prote
22	82.5	60.7	233	2 T05153	hypothetical prote
23	82.5	60.7	568	2 S15008	gene disco protein
24	82	60.3	60	2 C64698	probable histidine
25	82	60.3	115	2 H72583	hypothetical prote
26	81.5	59.9	658	2 T04219	hypothetical prote
27	81.5	59.9	754	2 JC4898	Down-syndrome-crit
28	81	59.6	391	2 H86187	hypothetical prote
29	81	59.6	2649	2 T51023	hypothetical prote

30	80	58.8	102	2 T30119	hypothetical prote
31	80	58.8	149	2 A54530	eggshell protein -
32	80	58.8	351	2 T20270	hypothetical prote
33	79.5	58.5	311	2 A56235	transcription acti
34	79.5	58.5	950	2 S27473	URBS1 protein - gm
35	79	58.1	83	2 T16435	hypothetical prote
36	79	58.1	302	2 A55641	homeotic protein G
37	79	58.1	474	2 A40721	neuroblast prolife
38	79	58.1	657	2 A29454	knob-associated hi
39	79	58.1	735	2 T45059	hypothetical prote
40	79	58.1	1305	2 A40879	phospholipase C (E
41	79	58.1	1312	1 B40879	phospholipase C (E
42	78	57.4	306	2 T44684	hypothetical prote
43	78	57.4	355	2 S35345	otx1 protein - mou
44	78	57.4	355	2 I56547	homeodomain protei
45	78	57.4	469	2 I37451	HBFG-2 (HFK-2) pro

ALIGNMENTS

RESULT 1

G84783

Probable pectinesterase [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: G84783

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, C.;

Nature 402, 761-768, 1999

A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: G84783

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-407 <STO>

A:Cross-references: GB:AE002093; NID:g4415916; PIDN:AAD20147.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g36710

A:Map position: 2

Query Match 70.6%; Score 96; DB 2; Length 407;
Best Local Similarity 68.4%; Pred. No. 0.00046;
Matches 13; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 KKHKKHHKKHHKKHHKK 19

Db 51 KKHKKHHKKHHKKHHKKH 69

RESULT 2

A54523

histidine-rich protein - Plasmodium lophurae (fragment)

C:Species: Plasmodium lophurae

C>Date: 15-Oct-1994 #sequence_revision 15-Oct-1994 #text_change 20-Aug-1999

C:Accession: A54523

R;Irving, D.O.; Cross, G.A.M.; Feder, R.; Wallach, M.

Mol. Biochem. Parasitol. 18, 223-234, 1986

A>Title: Structure and organization of the histidine-rich protein gene of Plasmodium lo

A:Reference number: A54523; MUID:86174893; PMID:3007981

A:Accession: A54523

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-140 <IRV>

A:Cross-references: GB:M15317; NID:g160331; PIDN:AAA29616.1; PID:g552196

C:Superfamily: plasmodium histidine-rich protein

C:Keywords: tandem repeat

Query Match 67.6%; Score 92; DB 2; Length 140;

Best Local Similarity 75.0%; Pred. No. 0.00049;

Matches 12; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

W d Jan 21 11:27:54 2004

us-10-018-103a-6.rpr

QY 3 HHHKHHKHHKHHKHH 18
||| ||| ||| ||| |||
Db 10 HHHHHHAPHHHHHH 25

RESULT 3

KGZQHL
histidine-rich glycoprotein precursor - Plasmodium lophurae
C:Species: Plasmodium lophurae
C:Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 16-Jul-1999
C:Accession: A22692
R:Raveech, J.V.; Feder, R.; Pavlovic, A.; Blobel, G.
Nature 312, 616-620, 1984
A:Title: Primary structure and genomic organization of the histidine-rich protein of the
A:Reference number: A22692; MUID:85061618; PMID:6095114
A:Accession: A22692
A:Molecule type: DNA
A:Residues: 1-351 <RAV>
A:Cross-references: GB:X01469; NID:g9997; PIDN:CAA25698.1; PID:g9999
C:Comment: There are two copies of 16-residue repeats, two copies of 17-residue repeats,
C:Genetics:
A:Introns: 23/3
A:Superfamily: plasmodium histidine-rich protein
C:Keywords: glycoprotein; tandem repeat
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-47/Domain: propeptide #status predicted <PRO>
F:48-351/Product: histidine-rich glycoprotein #status predicted <MAT>
F:59-74, 75-90/Region: 16-residue repeats
F:91-107, 108-123/Region: 17-residue repeats
F:124-138, 139-153/Region: 15-residue repeats
F:173-301, 312-331/Region: 10-residue repeats
F:40/Binding site: carbohydrate (Aen) #status predicted

Query Match 67.6%; Score 92; DB 1; Length 351;
Best Local Similarity 75.0%; Pred. No. 0.0011;
Matches 12; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 3 HHHKHHKHHKHHKHH 18
||| ||| ||| ||| |||
Db 176 HHHHHHAPHHHHHH 191

RESULT 4

A84696
probable zinc transporter [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: A84696
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: A84696
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-385 <STO>
A:Cross-references: GB:AE002093; NID:g3980394; PIDN:AAC95197.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g29410
A:Map position: 2

Query Match 66.9%; Score 91; DB 2; Length 385;
Best Local Similarity 66.7%; Pred. No. 0.0016;
Matches 12; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 HHHKHHKHHKHHKHHK 20
||| ||| ||| ||| |||
Db 201 HHHHHHAKQHHK 218

RESULT 5

A29653
histidine-rich protein - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 09-Jun-2000
C:Accession: A29653
R:Lenstra, R.; d'Auriol, L.; Andrieu, B.; Le Bras, J.; Galibert, F.
Biochem. Biophys. Res. Commun. 146, 368-377, 1987
A:Title: Cloning and sequencing of Plasmodium falciparum DNA fragments containing repeat
A:Reference number: A90134; MUID:87270765; PMID:3038111
A:Accession: A29653
A:Molecule type: DNA
A:Residues: 1-82 <LEN>
A:Cross-references: GB:M17028; NID:g160339; PID:g160341

Query Match 64.7%; Score 88; DB 2; Length 82;
Best Local Similarity 46.2%; Pred. No. 0.00083;
Matches 12; Conservative 4; Mismatches 2; Indels 8; Gaps 1;

QY 2 KHHH-----KHHKHHKHHKHHK 19
:|||||:|||||:|||||:
Db 12 RHHLNLHLYRHHHHHHHHHHHHR 37

RESULT 6

T13389
hypothetical protein I15C2.10 - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13389
R:Salles, C.; Valenti, P.; Darlameiou, A.; Henderson, N.; Campbell, L.; Glover, D.
submitted to the EMBL Data Library, May 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: Z17665

A:Accession: T13389
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1291 <CAT>
A:Cross-references: EMBL:AL031581; NID:e1320978; PID:e1426292; PIDN:CAA20894.2
C:Genetics:
A:Cross-references: FlyBase:FBgn0020381
A:Map position: X
A:Introns: 238/3; 1225/1
A:Note: EG:115C2.10

Query Match 64.0%; Score 87; DB 2; Length 1291;
Best Local Similarity 61.1%; Pred. No. 0.012;
Matches 11; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 KHHKHHKHHKHHKHHK 18
:|||||:|||||:|||||:
Db 535 QHHHQHHHHHHHHHHH 552

RESULT 7

A56837
homeotic protein MOX2 - human
C:Species: Homo sapiens (man)
C:Date: 11-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 24-Sep-1999
C:Accession: A56837; S52168
R:Grigoriou, M.; Kastiraki, M.C.; Modi, W.S.; Theodorakis, K.; Mankoo, B.; Pachnis, V.
Genomics 26, 550-555, 1995
A:Title: Isolation of the human MOX2 homeobox gene and localization to chromosome 7p22
A:Reference number: A56837; MUID:95331791; PMID:7607679
A:Accession: A56837
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-303 <GRI>
A:Cross-references: EMBL:X82629; NID:g732790; PIDN:CAA57949.1; PID:g732791
R:Grigoriou, M.; Kastiraki, M.C.; Modi, W.; Theodorakis, K.; Mankoo, B.; Pachnis, V.,
submitted to the EMBL Data Library, November 1994
A:Description: Isolation of the human MOX2 homeobox gene and localization to chromosome
A:Reference number: S52168
A:Accession: S52168

Query Match 66.9%; Score 91; DB 2; Length 385;
Best Local Similarity 66.7%; Pred. No. 0.0016;
Matches 12; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 HHHKHHKHHKHHKHHK 20
||| ||| ||| ||| |||
Db 201 HHHHHHAKQHHK 218

RESULT 5

A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-7,'R',9-110,'R',112-122,'Q',124-157,'V',159-303 <GR2>
 A;Cross-references: EMBL:X82629
 C;Genetics:

A;Gene: GDB:MOX2
 A;Cross-references: GDB:I20703; OMIM:155970
 A;Map position: 3q13-3q13
 C;Superfamily: unassigned homeobox proteins; homeobox homology
 C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F;187-243/Domain: homeobox homology <HOX>

Query Match 63.2%; Score 86; DB 2; Length 303;
 Best Local Similarity 57.9%; Pred. No. 0.0044;
 Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 KHHKHHKHHKHHKHHK 20
 :||| ||| ||| ||| ||| :
 Db 63 QHHRGHHHHHHHHHHHQ 81

RESULT 8

B49122
 homeobox protein Mox-2 - mouse
 C;Species: Mus musculus (house mouse)
 C;Date: 18-Dec-1993 #sequence_revision 17-Nov-1995 #text_change 17-Nov-2000
 C;Accession: B49122; S41779; S29902
 R;Candia, A.F.; Hu, J.; Crosby, J.; Lalley, P.A.; Noden, D.; Nadeau, J.H.; Wright, C.V.
 Development 116, 1123-1136, 1992
 A;Title: Mox-1 and Mox-2 define a novel homeobox gene subfamily and are differentially expressed in the mouse embryo.
 A;Reference number: A49122; MUID:93201999; PMID:1363541
 A;Accession: B49122

A;Status: preliminary; nucleic acid sequence not shown
 A;Molecule type: mRNA
 A;Residues: 1-303 <CAN1>

A;Cross-references: EMBL:Z16406; NID:G57949; PIDN:CAA78899.1; PID:G57950
 A;Experimental source: C57BL
 A;Note: the complete translation is not shown
 R;Candia, A.F.; Kovalik, J.P.; Wright, C.V.E.
 Nucleic Acids Res. 21, 4982, 1993

A;Title: Amino acid sequence of Mox-2 and comparison to its Xenopus and rat homologs.
 A;Reference number: S41469; MUID:94232829; PMID:7909944
 A;Accession: S41779

A;Status: preliminary; nucleic acid sequence not shown
 A;Molecule type: mRNA
 A;Residues: 1-303 <CAN2>

A;Cross-references: EMBL:Z16406; NID:G57949; PIDN:CAA78899.1; PID:G57950
 R;Candia, A.
 submitted to the EMBL Data Library, October 1992

A;Reference number: S29902
 A;Accession: S29902

A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 186-246 <CAN>
 A;Cross-references: EMBL:Z16406
 C;Superfamily: unassigned homeobox proteins; homeobox homology
 C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F;187-243/Domain: homeobox homology <HOX>

Query Match 63.2%; Score 86; DB 2; Length 303;
 Best Local Similarity 57.9%; Pred. No. 0.0044;
 Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 KHHKHHKHHKHHKHHK 20
 :||| ||| ||| ||| ||| :
 Db 63 QHHRGHHHHHHHHHHHQ 81

RESULT 9

A48130
 growth arrest-specific homeobox protein Cvx - rat
 C;Species: Rattus norvegicus (Norway rat)
 C;Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 24-Sep-1999

C;Accession: A48130; S31976
 R;Gorski, D.H.; LePage, D.F.; Patel, C.V.; Copeland, N.G.; Jenkins, N.A.; Walsh, K.
 Mol. Cell. Biol. 13, 3722-3733, 1993
 A;Title: Molecular cloning of a diverged homeobox gene that is rapidly down-regulated d
 A;Reference number: A48130; MUID:93268321; PMID:8098844
 A;Accession: A48130

A;Status: preliminary
 A;Molecule type: nucleic acid
 A;Residues: 1-303 <GOR>
 A;Cross-references: GB:Z17223; NID:G57951; PIDN:CAA78931.1; PID:G57952
 A;Experimental source: aorta

A;Note: Sequence extracted from NCBI backbone (NCBIN:132842, NCBIP:132843)
 R;Gorski, D.H.; LePage, D.F.; Patel, C.V.; Copeland, N.G.; Jenkins, N.A.; Walsh, K.
 submitted to the EMBL Data Library, October 1992

A;Description: Cvx: A diverged homeodomain gene which is rapidly down-regulated followi
 A;Reference number: S31976

A;Accession: S31976

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-182,'S',184-303 <GO2>

A;Cross-references: EMBL:Z17223

C;Superfamily: unassigned homeobox proteins; homeobox homology

C;Keywords: DNA binding; homeobox; nucleus; transcription regulation

F;187-243/Domain: homeobox homology <HOX>

Query Match 63.2%; Score 86; DB 2; Length 303;
 Best Local Similarity 57.9%; Pred. No. 0.0044;
 Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 KHHKHHKHHKHHKHHK 20
 :||| ||| ||| ||| ||| :
 Db 63 QHHRGHHHHHHHHHHHQ 81

RESULT 10

S66671

neuron-derived receptor NOR-2 - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 20-Sep-1999

C;Accession: S66671

R;Retopoulos, I.; Part, D.; Ochoa, A.; Zakin, M.M.; Lamas, E.

FEBS Lett. 372, 273-278, 1995

A;Title: NOR-2 (neuron-derived orphan receptor), a brain zinc finger protein, is highly

A;Reference number: S66671; MUID:96000221; PMID:7556683

A;Accession: S66671

A;Status: translation not shown

A;Molecule type: mRNA

A;Residues: 1-430 <PET>

A;Cross-references: EMBL:X86003; NID:G1065885; PIDN:CAA59993.1; PID:G1065886

C;Genetics:

A;Gene: nor-2

C;Superfamily: unassigned erba-related proteins; erba transforming protein homology
 C;Keywords: DNA binding; zinc finger

F;292-430/Domain: erba transforming protein homology #status atypical <ERBA>

F;292-379/Region: DNA binding

F;294-314/Region: zinc finger CCCC motif

F;312-316/Region: P box

F;330-354/Region: zinc finger CCCC motif

F;331-335/Region: D box

F;372-378/Region: A box

Query Match 62.5%; Score 85; DB 2; Length 430;
 Best Local Similarity 66.7%; Pred. No. 0.0078;
 Matches 12; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

QY 3 HHHKHHKHHKHHKHHK 20
 :||| ||| ||| ||| :
 Db 99 HHHHHH--HHHHHHHQ 114

RESULT 11

S71930
 neuron-derived receptor NOR-1 - human

C;Species: Homo sapiens (man)
C;Date: 04-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 20-Feb-1998
C;Accession: S71930
R;Ohkura, N.; Ito, M.; Tsukada, T.; Sasaki, K.; Yamaguchi, K.; Miki, K.
Biochim. Biophys. Acta 1308, 205-214, 1996
A;Title: Structure, mapping and expression of a human NOR-1 gene, the third member of the
A;Reference number: S71930; MUID:96404972; PMID:8809112
A;Accession: S71930
A;Molecule type: mRNA
A;Residues: 1-625 <OHK>
A;Cross-references: EMBL:D78579
A;Note: DNA was also sequenced
C;Genetics:
A;Map position: 9q
C;Superfamily: Probable hormone receptor N10, nuclear; erBA transforming protein homolog
F;230-544/Domain: erBA transforming protein homology <ERBA>
F;290-377/Domain: DNA binding #status predicted <DNA>
F;292-312/Region: zinc finger CCCC motif
F;310-314/Region: P box
F;328-352/Region: zinc finger CCCC motif
F;329-333/Region: D box
F;370-376/Region: A box

Query Match 62.5%; Score 85; DB 2; Length 625;
Best Local Similarity 66.7%; Pred. No. 0.011; Indels 2; Gaps 1;
Matches 12; Conservative 2; Mismatches 2

QY 3 HHHKHHKHHKHHKHHK 20
DB 95 HHHHHH--HHHHHHQ 110

RESULT 12
JC2493
neuron derived orphan receptor-1 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 16-Mar-1995 #sequence_revision 26-May-1995 #text_change 20-Jun-2000
C;Accession: JC2493; I48179
R;Ohkura, N.; Hijiuro, M.; Yamamoto, A.; Miki, K.
Biochem. Biophys. Res. Commun. 205, 1959-1965, 1994
A;Title: Molecular cloning of a novel thyroid/steroid receptor superfamily gene from cul
A;Reference number: JC2493; MUID:95110348; PMID:7811288
A;Accession: JC2493
A;Molecule type: mRNA
A;Residues: 1-628 <OHK>
A;Cross-references: DDBJ:D38530; NID:g1483194; PIDN:BA07535.1; PID:g643600
A;Experimental source: neuronal cell
R;Pena de Ortiz, S.; Cannon, M.M.; Jamieson, G.A.
Brain Res. Mol. Brain Res. 23, 278-283, 1994
A;Title: Expression of nuclear hormone receptors within the rat hippocampus: identificat
A;Reference number: I48179; MUID:94335560; PMID:7914660
A;Accession: I48179
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 318-345 <RES>
A;Cross-references: GB:I19343; NID:g450942; PIDN:AA046395.1; PID:g565379
A;Experimental source: hippocampus
C;Superfamily: probable hormone receptor N10, nuclear; erBA transforming protein homolog
C;Keywords: DNA binding; receptor; zinc finger
F;292-547/Domain: erBA transforming protein homology <ERBA>
F;292-379/Region: DNA binding #status predicted
F;312-316/Region: P box
F;331-335/Region: D box
F;372-378/Region: A box

Query Match 62.5%; Score 85; DB 2; Length 628;
Best Local Similarity 66.7%; Pred. No. 0.011; Indels 2; Gaps 1;
Matches 12; Conservative 2; Mismatches 2

QY 3 HHHKHHKHHKHHKHHK 20
DB 95 HHHHHH--HHHHHHQ 110

Db 99 HHHHHH--HHHHHHQ 114

RESULT 13
T13804
shs protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: T13804
R;Treisman, J.E.; Lai, Z.C.; Rubin, G.M.
Development 121, 2835-2845, 1995
A;Title: Shortighted acts in the decapentaplegic pathway in Drosophila eye development
A;Reference number: Z17767; MUID:96038094; PMID:7555710
A;Accession: T13804
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1212 <TRE>
A;Cross-references: EMBL:L42512; NID:g833709; PID:g833710; PIDN:AA041608.1
C;Genetics:
A;Gene: shs
A;Cross-references: FlyBase:FBgn0010460

Query Match 62.5%; Score 85; DB 2; Length 1212;
Best Local Similarity 64.7%; Pred. No. 0.02; Indels 4; Gaps 0;
Matches 11; Conservative 2; Mismatches 4

QY 3 HHHKHHKHHKHHKHHK 19
DB 301 HHHQPHHHHHHHHHQ 317

RESULT 14
D83483
probable metal transporter PA1297 [imported] - Pseudomonas aeruginosa (strain PA01)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 06-Jan-2003
C;Accession: D83483
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim.
Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: D83483
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-326 <STO>
A;Cross-references: GB:AE004559; GB:AE004091; NID:g9947228; PIDN:AA04686.1; GSPDB:GN00
A;Experimental source: strain PA01
C;Genetics:
A;Gene: PA1297
C;Superfamily: zinc transporter Znf-2

Query Match 61.8%; Score 84; DB 2; Length 326;
Best Local Similarity 73.3%; Pred. No. 0.0078; Indels 4; Gaps 0;
Matches 11; Conservative 0; Mismatches 4

QY 3 HHHKHHKHHKHHKHHK 17
DB 164 HHHHHHHHHHHHHH 178

RESULT 15
A38565
polycomb (Pc) protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A38565
R;Paro, R.; Hogness, D.S.
Proc. Natl. Acad. Sci. U.S.A. 88, 263-267, 1991
A;Title: The polycomb protein shares a homologous domain with a heterochromatin-associ
A;Reference number: A38565; MUID:91095442; PMID:1898775
A;Accession: A38565

Search completed: January 20, 2004, 18:24:44
Job time : 10.1449 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 18:06:29 ; Search time 6.66667 Seconds
(without alignments)
141.080 Million cell updates/sec

Title: US-10-018-103A-6
Perfect score: 136
Sequence: 1 KKHKKHHKHHKHHKHHK 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	92	67.6	351	1 HRPX_PLALO	P04929 plasmodium
2	92	67.6	428	1 FXB2_MOUSE	Q64733 mus musculus
3	88	64.7	82	1 HRP3_PLAFS	P14586 plasmodium
4	87	64.0	465	1 HNF6_HUMAN	Q9ubc0 homo sapien
5	87	64.0	465	1 HNF6_MOUSE	O08755 mus musculus
6	87	64.0	465	1 HNF6_RAT	P70512 rattus norv
7	86.5	63.6	485	1 ONC2_HUMAN	Q95948 homo sapien
8	86	63.2	303	1 MOX2_HUMAN	P50222 homo sapien
9	86	63.2	303	1 MOX2_MOUSE	P32443 mus musculus
10	86	63.2	303	1 MOX2_RAT	P39020 rattus norv
11	85	62.5	430	1 NOR2_RAT	Q63516 rattus norv
12	85	62.5	626	1 NR43_HUMAN	Q92570 homo sapien
13	85	62.5	628	1 NR43_RAT	P51179 rattus norv
14	85	62.5	1211	1 BUN2_DROME	Q24523 drosophila
15	84	61.8	390	1 PC_DROME	P26017 drosophila
16	83	61.0	420	1 YBE1_SCHPO	Q42980 schizosacch
17	83	61.0	469	1 A2AC_DIDMA	P35405 didelphis m
18	82.5	60.7	143	1 Y970_METJA	Q58380 methanococ
19	82.5	60.7	147	1 YV59_CAEEL	P50439 caenorhabdi
20	82.5	60.7	568	1 DISC_DROME	P23792 drosophila
21	82.5	60.7	1321	1 PGCN_HUMAN	O14594 homo sapien
22	82	60.3	59	1 HPN_HELPY	Q48251 helicobacte
23	82	60.3	558	1 CBX4_HUMAN	O00257 homo sapien
24	82	60.3	2084	1 CHDB_HUMAN	Q9hck8 homo sapien
25	81.5	59.9	763	1 DYRA_HUMAN	Q13627 homo sapien
26	81.5	59.9	763	1 DYRA_MOUSE	Q61214 mus musculus
27	81.5	59.9	763	1 DYRA_RAT	P63470 rattus norv
28	80	58.8	149	1 EGGS_SCHMA	P08016 schistosoma
29	80	58.8	351	1 CAV2_CAEEL	Q18879 caenorhabdi
30	80	58.8	977	1 DLP3_RAT	P97838 rattus norv
31	79.5	58.5	950	1 URBI_USTMA	P40349 ustilago ma
32	79	58.1	437	1 ZP12_BRARE	P56224 brachydanio
33	79	58.1	441	1 P033_BRARE	Q90436 brachydanio

34 79 58.1 442 1 VATC_DROME
35 79 58.1 474 1 ANA_DROME
36 79 58.1 620 1 SNF1_CANAL
37 79 58.1 657 1 KNOB_PLAFN
38 79 58.1 1312 1 PIP1_DROME
39 78 57.4 355 1 OTX1_MOUSE
40 78 57.4 355 1 OTX1_RAT
41 78 57.4 469 1 FXGA_HUMAN
42 78 57.4 477 1 FXGB_HUMAN
43 78 57.4 480 1 FXGB_RAT
44 78 57.4 481 1 FXGB_MOUSE
45 77 56.6 307 1 YK35_YEAST

Q9v7n5 drosophila
Q26307 drosophila
P52497 candida alb
P06719 plasmodium
P25455 drosophila
P80205 mus musculus
Q63410 rattus norv
P55316 homo sapien
P55315 homo sapien
Q00939 rattus norv
Q60987 mus musculus
P36155 saccharomyc

ALIGNMENTS

RESULT 1
ID HRPX_PLALO STANDARD; PRT; 351 AA.
AC P04929;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histidine-rich glycoprotein precursor.
OS Plasmodium lophurae.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID:5853;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85061618; PubMed=6095114;
RA Ravetch J.V., Feder R., Pavlovic A., Blobel G.;
RT "Primary structure and genomic organization of the histidine-rich
RT protein of the malaria parasite Plasmodium lophurae.";
RL Nature 312:616-620(1984).
CC -!- MISCELLANEOUS: IN THE INTRAERYTHROCYTIC STAGES OF DEVELOPMENT OF
CC P. LOPHURAE IN DUCKS, THERE IS A SYNTHESIS OF A MAJOR PROTEIN THAT
CC ACCUMULATES TO COMPRISE AT LEAST 50% OF THE CELLULAR MASS: THE
CC HISTIDINE RICH PROTEIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X01469; CAA25698.1; -
CC PIR; A22692; KGZQHL.
CC Malaria; Repeat; Glycoprotein; Signal.
CC SIGNAL 1 23
CC PROPEP 24 47
CC CHAIN 48 351
CC CARBOHYD 40 40
CC DOMA 59 90
CC REPEAT 59 74
CC REPEAT 75 90
CC REPEAT 91 123
CC DOMAIN 91 123
CC REPEAT 91 107
CC REPEAT 108 123
CC REPEAT 124 153
CC DOMAIN 124 138
CC REPEAT 139 153
CC REPEAT 173 351
CC SQ SEQUENCE 351 AA; 44032 MW; D19A48D47D890453 CRC64;
CC -----

HISTIDINE-RICH GLYCOPROTEIN.
N-LINKED (GLCNAC. . .) (PROBABLE).
2 X 16 AA TANDEM REPEATS.
16-1.
16-2.
2 X 17 AA TANDEM REPEATS.
17-1.
17-2.
2 X 15 AA TANDEM REPEATS.
15-1.
15-2.
18 X 10 AA TANDEM REPEATS.
D19A48D47D890453 CRC64;

Query Match 67.6%; Score 92; DB 1; Length 351;

Best Local Similarity 75.0%; Pred. No. 0.00097;
Matches 12; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 HHHKHHKHHKHHKHHH 18
||| ||| ||| ||| |||

```
Db 176 HHHHHHHHHHHHHH 191
RESULT 2
FXB2_MOUSE STANDARD; PRT; 428 AA.
ID -FXB2_MOUSE
AC Q64733;
DT 01-NOV-1997 (Rel. 35, Last Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Forkhead box protein B2 (Transcription factor FXH-4).
GN FOXB2 OR FXH4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCB1_TaxID=10090;
OX [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=97014266; PubMed=8861101;
RX Kaestner K.H., Schuetz G., Monaghan A.P.;
RA "Expression of the winged helix genes fkh-4 and fkh-5 defines domains
RT in the central nervous system.";
RT Mech. Dev. 55:221-230 (1996).
[2]
RN SEQUENCE OF 4-114 FROM N.A.
RP STRAIN=129;
RX MEDLINE=93361500; PubMed=7689224;
RA Kaestner K.H., Lee K.H., Schloendorff J., Hiemisch H.,
RA Monaghan A.P., Schuetz G.;
RA "Six members of the mouse forkhead gene family are developmentally
RT regulated.";
RT Proc. Natl. Acad. Sci. U.S.A. 90:7628-7631 (1993).
RL Proc. CELLULAR LOCATION: Nuclear.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED DURING EMBRYOGENESIS.
CC -!- SIMILARITY: Contains 1 fork-head domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X92591; CAA63335.1; -
DR EMBL; X71942; CAA50744.1; -
DR PIR; D47746; D47746.
DR HSP; G63245; 2HFH.
DR TRANSFAC; T02442; -
DR MGD; MGI:1347468; Foxb2.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS50039; FORK_HEAD_3; 1.
KW DNA-binding; Nuclear protein; Transcription regulation.
FT DNA BIND 12 103
FT DOMAIN 139 153 POLY-HIS.
FT DOMAIN 156 162 POLY-HIS.
FT DOMAIN 163 172 POLY-PRO.
FT DOMAIN 217 231 POLY-ALA.
FT DOMAIN 249 258 POLY-ALA.
FT DOMAIN 321 330 POLY-ALA.
FT DOMAIN 396 399 POLY-ALA.
FT SEQUENCE 428 AA; 45170 MW; DB8A8EFD1E94AB10 CRC64;
Query Match 67.6%; Score 92; DB 1; Length 428;
Best Local Similarity 75.08; Pred. No. 0.0012;
Matches 12; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 HHHHHHHHHHHHHH 18
Db 147 HHHHHHHHHHHHHH 162
RESULT 3
HRP3_PLAFS STANDARD; PRT; 82 AA.
ID -HRP3_PLAFS
AC P14586;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-MAY-1991 (Rel. 18, Last annotation update)
DE Histidine-rich protein.
OS Plasmodium falciparum (isolate fcm17 / Senegal).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OC NCB1_TaxID=5845;
OX [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=87270765; PubMed=3038111;
RX Lenstra R., D'Auriol L., Andrieu B., le Bras J., Galibert F.;
RA "Cloning and sequencing of Plasmodium falciparum DNA fragments
RT containing repetitive regions potentially coding for histidine-rich
RT proteins: identification of two overlapping reading frames.";
RL Biochem. Biophys. Res. Commun. 146:368-377 (1987)
CC -!- MISCELLANEOUS: THIS PROTEIN IS CODED ON THE REVERSE STRAND OF AN
CC ASPARTIC ACID-RICH PROTEIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M17028; AAA29619.1; -
DR PIR; A29653; A29653.
DR Malaria.
SQ SEQUENCE 82 AA; 10578 MW; 7CE2EA69F2FC1E8C CRC64;
Query Match 64.7%; Score 88; DB 1; Length 82;
Best Local Similarity 46.2%; Pred. No. 0.00067;
Matches 12; Conservative 4; Mismatches 2; Indels 8; Gaps 1;

QY 2 KHHH-----KHHHKKHHKHHK 19
Db 12 RHHLNHLHYRHHHHHHHHHHH 37
RESULT 4
HNF6_HUMAN STANDARD; PRT; 465 AA.
ID -HNF6_HUMAN
AC Q9UBC0; Q99744; Q9UMR6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hepatocyte nuclear factor 6 (HNF-6) (One cut domain family member 1).
GN ONECUT1 OR HNF6A OR HNF6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCB1_TaxID=9606;
OX [1]
RN SEQUENCE FROM N.A.
RP Furuta H., Horikawa Y., Wang Y.-Q., Bell G.I.;
RA "Isolation and characterization of the human hepatocyte nuclear factor
RT 6 gene.";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP Furuta H., Wang Y.-Q., Bell G.I.;
RA "The sequence of human mRNA for the hepatocyte nuclear factor-6
RT alpha.";
```


Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 4 HHKHHKHHKHHKHHK 20
DB 124 HHHHHHHHHHPHHQR 140

RESULT 6
HNF6 RAT
ID -HNF6_RAT STANDARD; PRT; 465 AA.
AC P70512; 088755;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hepatocyte nuclear factor 6 (HNF-6) (One cut domain family member 1).
GN ONECUT1 OR HNF6A OR HNF6.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM ALPHA).
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=96382488; PubMed=8790352;
RA Lemaigre F.P., Durvieux S.M., Truong O., Lannoy V.J., Hauan J.J.,
Rousseau G.G.;
RT "Hepatocyte nuclear factor 6, a transcription factor that contains a
novel type of homeodomain and a single cut domain."
RL Proc. Natl. Acad. Sci. U.S.A. 93:9460-9464 (1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM BETA).
RC STRAIN=Wistar;
RX MEDLINE=98256275; PubMed=9593691;
RA Lannoy V.J., Bueglin T.R., Rousseau G.G., Lemaigre F.P.;
RT "Isoforms of hepatocyte nuclear factor-6 differ in DNA-binding
properties, contain a bifunctional homeodomain, and define the new
ONECUT class of homeodomain proteins."
RL J. Biol. Chem. 273:13552-13562 (1998).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=97042457; PubMed=8887657;
RA Samadani U., Costa R.H.;
RT "The transcriptional activator hepatocyte nuclear factor 6 regulates
liver gene expression."
RL Mol. Cell. Biol. 16:6273-6284 (1996).
CC -!- FUNCTION: TRANSCRIPTIONAL ACTIVATOR. BINDS THE CONSENSUS SEQUENCE
5'-DHATTGATTTWD-3' ON A VARIETY OF GENE PROMOTERS SUCH AS THOSE
OF HNF3B AND TTR. IMPORTANT FOR LIVER GENES TRANSCRIPTION. THE
AFFINITY OF HNF-GALPHA AND HNF-6BETA FOR DNA DIFFERS DEPENDING ON
THE TARGET SEQUENCE.
CC -!- SUBUNIT: Binds DNA as a monomer.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Alpha;
CC IsoId=P70512-1; Sequence=Displayed;
CC Name=Beta;
CC IsoId=P70512-2; Sequence=VSP_002312;
CC TISSUE SPECIFICITY: EXPRESSED IN LIVER, BRAIN, SPLEEN AND TESTIS.
CC -!- SIMILARITY: BELONGS TO CUT DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE CUT HOMEBOX FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; Y96553; CAA65389.1; -;
CC EMBL; Y14933; CAA75150.1; -;
CC DR EMBL; Y14933; CAA75150.1; -;
CC DR TRANSFAC; T03257; -;

TRANSFAC; T03258; -;
DR InterPro; IPR007108; Cut_homeo.
DR InterPro; IPR003350; HmoEO CUT.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF02376; CUT; 1.
DR Pfam; PF00046; homeobox; 1.
DR ProDom; PD00010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX 1; FALSE_NEG.
DR PROSITE; PS00071; HOMEBOX 2; 1.
KW Transcription regulation; Homeobox; DNA-binding; Nuclear protein;
KW Activator; Alternative splicing.
FT DNA_BIND 283 369 CUT.
FT DNA_BIND 385 444 HOMEBOX.
FT DOMAIN 69 72 POLY-HIS.
FT DOMAIN 124 138 POLY-HIS.
FT DOMAIN 455 460 POLY-SER.
FT VARSPLIC 368 368 A -> AESAMGVSVPSLRITSGGQLSVPLP (in
isoform Beta).
FT /FTid=VSP_002312.
FT
SQ SEQUENCE 465 AA; 51067 MW; BD651267FD7AC896 CRC64;
Query Match 64.0%; Score 87; DB 1; Length 465;
Best Local Similarity 64.7%; Pred. No. 0.0042;
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 4 HHKHHKHHKHHKHHK 20
DB 124 HHHHHHHHHHPHHQR 140
RESULT 7
ONC2 HUMAN STANDARD; PRT; 485 AA.
ID -ONC2_HUMAN
AC G95948;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE One cut domain family member 2 (ONECUT-2 transcription factor) (OC-2).
GN ONECUT2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99115605; PubMed=9915796;
RA Jacquemin P., Lannoy V., Rousseau G.G., Lemaigre F.P.;
RT "OC-2, a novel mammalian member of the ONECUT class of homeodomain
transcription factors whose function in liver partially overlaps with
that of hepatocyte nuclear factor-6."
RL J. Biol. Chem. 274:2665-2671 (1999).
CC -!- FUNCTION: TRANSCRIPTIONAL ACTIVATOR. ACTIVATES THE TRANSCRIPTION
OF A NUMBER OF LIVER GENES SUCH AS HNF3B.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Contains 1 CUT domain.
CC -!- SIMILARITY: BELONGS TO THE CUT HOMEBOX FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; Y18198; CAB38253.1; -;
DR TRANSFAC; T03259; -;
DR Genew; HGNC:8139; ONECUT2.
DR MIM; 604894; -;
DR GO; GO:0003702; F:RNA polymerase II transcription factor acti. . .; TAS.
DR GO; GO:0007397; P:histogenesis and organogenesis; TAS.
DR InterPro; IPR007108; Cut_homeo.

```
DR InterPro; IPR003350; Hmoec_CUT.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF02376; CUT; 1.
DR Pfam; PF00046; homeobox; 1.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX 1; FALSE_NEG.
DR PROSITE; PS50071; HOMEBOX 2; 1.
KW Transcription regulation; Homeobox; DNA-binding; Nuclear protein;
KW Activator.
FT DNA_BIND 305 391 CUT.
FT DNA_BIND 407 466 HOMEBOX.
FT DOMAIN 18 37 POLY-GLY.
FT DOMAIN 62 66 POLY-PRO.
FT DOMAIN 75 82 POLY-ALA.
FT DOMAIN 152 165 POLY-HIS.
FT DOMAIN 298 303 POLY-SER.
SQ SEQUENCE 485 AA; 52482 MW; AF21E052EFBE5DA1 CRC64;

Query Match 63.6%; Score 86.5; DB 1; Length 485;
Best Local Similarity 61.9%; Pred. No. 0.0049;
Matches 13; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

QY 1 KKH-HKHKKKKKKKKKKKK 20
DB 147 KFHHPHPPHHHHHHHHQ 167

RESULT 8
MOX2_HUMAN STANDARD; PRT; 303 AA.
AC P50222; Q9UPL6;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Homeobox protein MOX2 (Mesenchyme homeobox 2) (Growth arrest-specific
DE homeobox).
GN ME0X2 OR MOX2 OR GAX.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=Embryo;
RC MEDLINE=95331791; PubMed=7607679;
RA Grigoriou M., Kastiraki M.-C., Modi W., Theodorakis K., Mankoo B.,
RA Pachnis V., Karagozeos D.;
RT "Isolation of the human MOX2 homeobox gene and localization to
RT chromosome 7p22.1-p21.3.";
RL Genomics 26:550-555(1995).
[2]
RN SEQUENCE FROM N.A.
RP TISSUE=Heart;
RC MEDLINE=95229154; PubMed=7713505;
RA Lepage D.F., Walsh K.;
RT "Molecular cloning and localization of the human GAX gene to 7p21.";
RL Genomics 24:535-540(1994).
[3]
RN SEQUENCE FROM N.A.
RP TISSUE=Skin;
RC MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano P.J., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McSwan N.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
```

Wed Jan 21 11:27:54 2004

Query Match 63.2%; Score 86; DB 1; Length 303;
 Best Local Similarity 57.9%; Pred. No. 0.0036;
 Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 2 KHHKHHKHHKHHKHHK 20
 DB 63 QHRRGHHHHHHHHHHHQ 81

RESULT 10

MOX2 RAT STANDARD; PRT; 303 AA.

ID MOX2 RAT AC P39020;

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 01-FEB-1995 (Rel. 31, Last annotation update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Homeobox protein MOX-2 (Growth arrest-specific homeobox).

GN ME0X2 OR MOX2 OR MOX-2 OR GAX.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1] SEQUENCE FROM N.A.

RP TISSUE=Aorta;

RC MEDLINE=93268321; PubMed=8098844;

RA Gorski D.H., Lepage D.F., Patel C.V., Copeland N.G., Jenkins N.A.,

RA Walsh K.;

RT "Molecular cloning of a diverged homeobox gene that is rapidly down-

regulated during the G0/G1 transition in vascular smooth muscle

cells.";

RT Mol. Cell. Biol. 13:3722-3733(1993).

RN [2] REVISIONS.

RP Walsh K.;

RA Submitted (MAR-1993) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: ROLE IN MESODERM INDUCTION AND ITS EARLIEST REGIONAL

SPECIFICATION, SOMITOGENESIS, AND MYOGENIC AND SCLEROTOMAL

DIFFERENTIATION. MAY HAVE A REGULATORY ROLE WHEN QUIESCENT

CC VASCULAR SMOOTH MUSCLE CELLS REENTER THE CELL CYCLE.

CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).

CC -1- TISSUE SPECIFICITY: AORTA AND HEART. ALSO DETECTED IN LUNG AND

KIDNEY.

CC -1- INDUCTION: RAPIDLY AND TRANSIENTLY DOWN-REGULATED DURING THE

TRANSITION FROM G0 TO G1 INDUCED BY MITOGEN STIMULATION.

CC -1- SIMILARITY: Contains 1 homeobox domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

EMBL; Z17223; CAA78931.1; -

PIR; A48130; A48130.

HSSP; P14653; 1B72.

TRANSFAC; T04048; -

InterPro; IPR001356; Homeobox.

Pfam; PF000046; homeobox; 1.

PRINTS; PR00031; HTHREPRESSR.

ProDom; PD000010; Homeobox; 1.

SMART; SM00389; HOX; 1.

PROSITE; PS00027; HOMEBOX 1; 1.

PROSITE; PS00071; HOMEBOX 2; 1.

Homeobox; DNA-binding; Nuclear protein; Developmental protein.

DOMAIN 42 47 POLY-SER.

DOMAIN 68 79 POLY-HIS.

DOMAIN 80 85 POLY-GLN.

DOMAIN 64 85 GLN/HIS-RICH (OPA-REPEAT).

DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DE Homeobox protein MOX-2 (Mesenchyme homeobox 2).
 GN ME0X2 OR MOX2 OR MOX-2 OR GAX.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1] SEQUENCE FROM N.A.

RP MEDLINE=93201999; PubMed=1363541;

RA Candia A.F., Hu J., Crosby J., Lallely P.A., Noden D., Nadeau J.H.,

RA Wright C.V.E.;

RT "Mox-1 and Mox-2 define a novel homeobox gene subfamily and are

differentially expressed during early mesodermal patterning in mouse

embryos.";

RT Development 116:1123-1136(1992).

RN [2] SEQUENCE FROM N.A.

RP MEDLINE=94232829; PubMed=7909944;

RA Candia A.F., Kovalik J.-P., Wright C.V.E.;

RA "Amino acid sequence of Mox-2 and comparison to its Xenopus and rat

homologs.";

RT Nucleic Acids Res. 21:4982-4982(1993).

RN [3] SEQUENCE OF 1-11 FROM N.A.

RP MEDLINE=95349593; PubMed=7623821;

RA Andres V., Fisher S., Weatsch P., Walsh K.;

RA "Regulation of Gax homeobox gene transcription by a combination of

positive factors including myocyte-specific enhancer factor 2.";

RT Mol. Cell. Biol. 15:4272-4281(1995).

CC -1- FUNCTION: Role in mesoderm induction and its earliest regional

specification, somitogenesis, and myogenic and sclerotomal

differentiation. May have a regulatory role when quiescent

CC vascular smooth muscle cells reenter the cell cycle.

CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).

CC -1- DEVELOPMENTAL STAGE: It is not expressed before 8-8.5 dpc. At 8-

8.5 dpc it is found on the entire epithelium of the somite. At 9.5

dpc its expression is restricted to the sclerotome. At 10.5 dpc it

is found in sclerotomally derived cells including the vertebral

and costal precursors.

CC -1- SIMILARITY: Contains 1 homeobox domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

EMBL; Z16406; CAA78899.1; -

PIR; B49122; B49122.

HSSP; P14653; 1B72.

TRANSFAC; T04048; -

MGI; MGI:103219; Meox2.

InterPro; IPR001356; Homeobox.

Pfam; PF000046; homeobox; 1.

PRINTS; PR00031; HTHREPRESSR.

ProDom; PD000010; Homeobox; 1.

SMART; SM00389; HOX; 1.

PROSITE; PS00027; HOMEBOX 1; 1.

PROSITE; PS00071; HOMEBOX 2; 1.

Homeobox; DNA-binding; Nuclear protein; Developmental protein.

DOMAIN 42 47 POLY-SER.

DOMAIN 68 79 POLY-HIS.

DOMAIN 80 85 POLY-GLN.

DOMAIN 186 245 GLN/HIS-RICH (OPA-REPEAT).

DOMAIN BIND 303 AA; 33506 MW; 41BD05FC39AA4427 CRC64;

SEQUENCE

FT DNA_BIND 186 245 HOMEBOX.
SQ SEQUENCE 303 AA; 33605 MW; 7776642AEFA3A2E8 CRC64;
Query Match 63.2%; Score 86; DB 1; Length 303;
Best Local Similarity 57.9%; Pred. No. 0.0036;
Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 KHHKHHKHHKHHKHHK 20
DB 63 QHRRGHHHHHHHHHHQ 81

RESULT 11
NOR2_RAT
ID NOR2_RAT STANDARD; PRT; 430 AA.
AC Q63516;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Nuclear hormone receptor NOR-2 (Neuron-derived orphan receptor 2).
GN NOR2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=9600221; PubMed=7556683;
RA Petropoulos I., Part D., Ochoa A., Zakim M.M., Lamas E.;
RT "NOR-2 (neuron-derived orphan receptor), a brain zinc finger protein,
RT is highly induced during liver regeneration.";
RL PBES Lett. 372:273-278(1995).
CC -1- FUNCTION: BINDS TO NRE.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- TISSUE SPECIFICITY: FOUND IN ALL TISSUES TESTED.
CC -1- SIMILARITY: Belongs to the nuclear hormone receptor family. NR4
subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X86003; CAA59993.1; -.
DR PIR; S66671; S66671.
DR HSSP; P19793; 2NLL.
DR TRANSFAC; T04752; -.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00047; STROIDFINGER.
DR ProDom; PD000035; Znf_C4steroid; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger.
FT DOMAIN 99 112 POLY-HIS.
FT DOMAIN 113 116 POLY-GLN.
FT DOMAIN 186 189 POLY-ALA.
FT DOMAIN 219 222 POLY-ALA.
FT DOMAIN 229 233 POLY-ALA.
FT DOMAIN 286 289 POLY-SER.
FT DNA_BIND 294 359 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 294 314 C4-TYPE.
FT ZN_FING 330 354 C4-TYPE.
SQ SEQUENCE 430 AA; 46297 MW; A30842899EF60FFB CRC64;
Query Match 62.5%; Score 85; DB 1; Length 430;
Best Local Similarity 56.7%; Pred. No. 0.0064;
Matches 12; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

QY 3 HHHKHHKHHKHHKHHK 20
DB 99 HHHHHH--HHHHHHQ 114

RESULT 12
NR43_HUMAN
ID NR43_HUMAN STANDARD; PRT; 626 AA.
AC Q92570; Q12935; Q14979; Q16420; Q9UEK2; Q9UEK3;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nuclear hormone receptor NOR-1 (Neuron-derived orphan receptor 1)
DE (Mitogen induced nuclear orphan receptor).
GN NR4A3 OR NOR1 OR MINOR OR CHN OR CSMP.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM ALPHA).
RX TISSUE=Fetal brain;
RX MEDLINE=96404972; PubMed=8809112;
RA Ohkura N., Ito M., Tsukada T., Sasaki K., Yamaguchi K., Miki K.;
RT "Structure, mapping and expression of a human NOR-1 gene, the third
RT member of the Nur77/NGFI-B family.";
RL Biochim. Biophys. Acta 1308:205-214(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM ALPHA).
RX TISSUE=Peripheal blood;
RX MEDLINE=96192925; PubMed=8614405;
RA Hedvat C.V., Irving S.G.;
RT "The isolation and characterization of MINOR, a novel mitogen-
RT inducible nuclear orphan receptor.";
RL Mol. Endocrinol. 9:1692-1700(1995).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM ALPHA).
RX TISSUE=Fetal brain;
RX MEDLINE=96152889; PubMed=8570200;
RA Clark J., Benjamin H., Gill S., Sidhar S., Goodwin G., Crew J.,
RA Gusterson B.A., Shipley J., Cooper C.S.;
RT "Fusion of the EWS gene to CHN, a member of the steroid/thyroid
RT receptor gene superfamily, in a human myxoid chondrosarcoma.";
RL Oncogene 12:223-235(1996).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM BETA), AND CHROMOSOMAL TRANSLOCATION WITH
RWS.
RX TISSUE=Fetal heart;
RX MEDLINE=96177652; PubMed=8634690;
RA Labelle Y., Zucman J., Stenman G., Kindblom L.-G., Knight J.,
RA Turc-Carel C., Dockhorn-Dworniczak B., Mandahl N., Desmaziere C.,
RA Peter M., Aurias A., Delattre O., Thomas G.;
RT "Oncogenic conversion of a novel orphan nuclear receptor by chromosome
RT translocation.";
RL Hum. Mol. Genet. 4:2219-2226(1995).
RN [5]
RP SEQUENCE OF 1-69 AND 301-443 FROM N.A. (ISOFORM BETA), AND ALTERNATIVE
RP SPLICING.
RX TISSUE=Skeletal muscle;
RX MEDLINE=98241504; PubMed=9573341;
RA Ohkura N., Ito M., Tsukada T., Sasaki K., Yamaguchi K., Miki K.;
RT "Alternative splicing generates isoforms of human neuron-derived
RT orphan receptor-1 (NOR-1) mRNA.";
RL Gene 211:79-85(1998).
CC -1- FUNCTION: BINDS TO THE B1A RESPONSE-ELEMENT.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Alpha;
CC IsoId=Q92570-1; Sequence=Displayed;
CC Name=Beta;
CC IsoId=Q92570-2; Sequence=VSP_003712, VSP_003713;

3 HHHHHHHKHHKHHKHHK 20
 ||| ||| ||| ||| ||| :
 95 HHHHHHH--HHHHHHHQ 110

Db

RESULT 13
 NR43 RAT
 ID NR43 RAT STANDARD; PRT; 628 AA.
 AC F51179;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Nuclear hormone receptor NOR-1 (Neuron-derived orphan receptor 1).
 GN NR4A3 OR NOR1
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 ON NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Wistar; TISSUE=Brain;
 RC MEDLINE=95110348; PubMed=7811288;
 RA Ohkura N., Hijikuro M., Yamamoto A., Miki K.;
 RT "Molecular cloning of a novel thyroid/steroid receptor superfamily
 gene from cultured rat neuronal cells.";
 RL Biochem. Biophys. Res. Commun. 205:1959-1965(1994).
 CC -1- FUNCTION: BINDS TO THE B1A RESPONSE-ELEMENT.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN CULTURED APOPTOTIC
 CC NEURONAL CELLS AND FETAL BRAIN, AND AT LOW LEVEL IN ADULT BRAIN.
 CC -1- SIMILARITY: Belongs to the nuclear hormone receptor family. NR4
 subfamily.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL; D38530; BAA07535.1; --
 DR PIR; JC2493; JC2493.
 DR HSSP; P19793; 2NLL.
 DR TRANSFAC; T02768; --
 DR InterPro; IPR000536; Hormone rec lig.
 DR InterPro; IPR001723; Stdhrmn_receptor.
 DR InterPro; IPR001628; ZnF_C4steroid.
 DR Pfam; PF00104; hormone_rec; 1.
 DR Pfam; PF00105; zf-C4; 1.
 DR PRINTS; PR00398; STRDHORMONER.
 DR PRINTS; PR00047; STROIDFINGER.
 DR ProDom; PD000035; ZnF_C4steroid; 1.
 DR SMART; SM00430; HOL1; 1.
 DR SMART; SM00339; ZnF_C4; 1.
 DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
 DR ProSite; Transcription regulation; DNA-binding; Nuclear protein;
 KW Zinc-finger.
 FT DOMAIN 99 112 POLY-HIS.
 FT DOMAIN 113 116 POLY-GLN.
 FT DOMAIN 186 189 POLY-ALA.
 FT DOMAIN 219 222 POLY-ALA.
 FT DOMAIN 229 233 POLY-ALA.
 FT DOMAIN 286 289 POLY-SER.
 FT DNA_BIND 294 359 NUCLEAR RECEPTOR-TYPE.
 FT ZN_FING 294 314 C4-TYPE.
 FT ZN_FING 330 354 C4-TYPE.
 FT DOMAIN 442 492 LIGAND-BINDING (POTENTIAL).
 SEQUENCE 628 AA; 68564 MW; 9AC285D9A65226D9 CRC64;

Query Match 62.5%; Score 85; DB 1; Length 628;
 Best Local Similarity 66.7%; Pred. No. 0.009; 2; Indels 2; Gaps 1
 Matches 12; Conservative 2; Mismatches

Oy 3 HHHHHHHHHHHHHHHHHHHK 20
Db 99 HHHHHHH--HHHHHHHQ 114

RESULT 14
BUN2_DROME STANDARD; PRT; 1211 AA.
ID BUN2_DROME Q24523; Q9VK78; 1211 AA.
AC Q24523; Q9VK78; 1211 AA.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Bunched protein, class 2/class 3 isoforms (Shortsighted protein).
GN BUN OR SHS OR CS461.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM CLASS 2), SUBCELLULAR LOCATION, AND
RP FUNCTION
RC TISSUE=Eye-antennal disk;
RX MEDLINE=96038094; PubMed=755710;
RA Treisman J.E., Lai Z.-C., Rubin G.M.;
RT "Shortsighted acts in the decapentaplegic pathway in Drosophila eye
RT development and has homology to a mouse TGF-beta-responsive gene";
RL Development 121:2835-2845(1995).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS CLASS 2 AND CLASS 3).
RC STRAIN=Berkely;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram D.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burkata K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fessler A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
CC -!- FUNCTION: Probable transcription factor required for peripheral
CC nervous system morphogenesis, eye development and oogenesis. May

be required for the transmission of the dpp signal and for a morphogenetic movement of the medulla in the brain that reorients the second optic lobe relative to the first. Plays a role in determining proper dorsal cell fates leading to the formation of the dorsal appendages. Nuclear and cytoplasmic.

-!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.

-!- ALTERNATIVE PRODUCTS:
Comment=Experimental confirmation may be lacking for some isoforms;
Name=Class 2;
IsoId=Q24523-1; Sequence=Displayed;
Name=Class 1;
IsoId=Q24522-1; Sequence=External;
Name=Class 3;
IsoId=Q24523-2; Sequence=VSP 006670;
-!- SIMILARITY: BELONGS TO THE TSC-22/DIP/BUN FAMILY.
-!- CAUTION: Ref.2 sequence differs from that shown due to erroneous gene model prediction.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; L42512; AAC1608.1; ...
EMBL; AB003636; AAF53200.1; ALT_SEQ.
EMBL; AB003636; AAF53201.1; ALT_SEQ.
DR F1R; T13804; T13804.
DR HSP; P80220; 1DIP.
DR FlyBase; FBgn0010460; bun.
DR InterPro; IPR000580; TSC-22_Dip_Bun.
DR Pfam; PF01166; TSC22; 1.
DR ProDom; PD007152; TSC-22_Dip_Bun; 1.
DR PROSITE; PS01289; TSC22; 1.
KW Transcription regulation; Nuclear protein; Alternative splicing.
FT DOMAIN 15 31 POLY-GLN.
FT DOMAIN 76 86 ALA-RICH.
FT DOMAIN 97 102 POLY-GLN.
FT DOMAIN 237 241 POLY-SER.
FT DOMAIN 249 254 POLY-GLN.
FT DOMAIN 261 265 POLY-SER.
FT DOMAIN 306 321 POLY-HIS.
FT DOMAIN 322 328 POLY-GLN.
FT DOMAIN 348 356 POLY-GLY.
FT DOMAIN 607 619 POLY-GLN.
FT DOMAIN 661 668 POLY-GLN.
FT DOMAIN 743 746 POLY-ALA.
FT DOMAIN 759 765 POLY-GLN.
FT DOMAIN 795 801 POLY-GLN.
FT DOMAIN 817 822 POLY-GLN.
FT DOMAIN 832 838 POLY-ALA.
FT DOMAIN 884 891 POLY-ALA.
FT DOMAIN 927 947 GLN-RICH.
FT DOMAIN 1001 1005 POLY-ALA.
FT DOMAIN 1011 1014 POLY-ALA.
FT DOMAIN 1069 1090 LEUCINE-ZIPPER.
FT DOMAIN 1194 1201 POLY-ALA.
FT VARSP LIC 1 109 Missing (in isoform Class 3).
FT FTID=VSP 006670.
FT K -> E (IN REF. 1).
FT CONFLICT 328 328 MISSING (IN REF. 1).
FT CONFLICT 801 801 Q -> QQQ (IN REF. 1).
FT CONFLICT 1189 1195 QQVTSAA -> TS (IN REF. 2).
SQ SEQUENCE 1211 AA; 125299 MW; 9925A9159A7051B0 CRC64;

Query Match 62.5%; Score 85; DB 1; Length 1211;
Best Local Similarity 64.7%; Pred. No. 0.016;
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Search completed: January 20, 2004, 18:23:56
Job time : 6.66667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 18:20:39 ; Search time 24.6377 Seconds
(without alignments)
209.478 Million cell updates/sec

Title: US-10-018-103A-6

Perfect score: 136

Sequence: 1 KXHHKHHKHHKHHKHHK 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	98	72.1	1318	5 Q95PH4	Q95ph4 dictyosteli
2	97	71.3	684	10 Q9SG87	Q9sg87 arabidopsis
3	96	70.6	407	10 Q92QA3	Q9zqa3 arabidopsis
4	94	69.1	404	13 Q8UVJ4	Q8uvj4 petromyzon
5	92	67.6	140	5 Q26056	Q26056 plasmodium
6	92	67.6	233	5 Q8MP30	Q8mp30 dictyosteli
7	92	67.6	1122	5 Q8SXD4	Q8sxd4 drosophila
8	91	66.9	109	11 Q9D6B9	Q9d6b9 mus musculus
9	91	66.9	385	10 Q9ZW23	Q9zw23 arabidopsis
10	89.5	65.8	4524	5 Q813J9	Q813j9 plasmodium
11	89	65.4	325	4 Q9BCN0	Q9bcn0 homo sapien
12	89	65.4	449	11 Q8CAJ8	Q8caj8 mus musculus
13	89	65.4	451	4 Q969F2	Q969f2 homo sapien
14	89	65.4	461	11 Q9D7U9	Q9d7u9 mus musculus
15	89	65.4	461	11 Q91Y45	Q91y45 mus musculus
16	89	65.4	461	11 Q8VE28	Q8ve28 mus musculus

17	89	65.4	678	5 Q94736	Q94736 stomoxys ca
18	88	64.7	109	5 Q8MXA9	Q8mx9 holopneuste
19	87	64.0	782	5 Q8T029	Q8t029 drosophila
20	87	64.0	825	5 Q8T3U6	Q8t3u6 drosophila
21	87	64.0	1281	13 Q918L4	Q918l4 brachydanio
22	87	64.0	1291	5 Q77261	Q77261 drosophila
23	87	64.0	1300	5 Q9W5E0	Q9w5e0 drosophila
24	87	64.0	2770	5 Q8ILV0	Q8ilv0 plasmodium
25	86	63.2	231	5 Q81QK7	Q81qk7 drosophila
26	86	63.2	302	13 Q90YH7	Q90yh7 gallus gall
27	86	63.2	303	6 Q95JA6	Q95ja6 aus scrofa
28	86	63.2	303	11 Q99M23	Q99m23 mus musculu
29	86	63.2	652	5 Q9VE69	Q9ve69 drosophila
30	86	63.2	725	5 Q87990	Q87990 drosophila
31	86	63.2	735	5 Q9W0L8	Q9w0l8 drosophila
32	86	63.2	837	5 Q9W4B2	Q9w4b2 drosophila
33	85	62.5	285	10 Q9LGP3	Q9lgp3 oryza sativ
34	85	62.5	1161	5 Q9W2X8	Q9w2x8 drosophila
35	85	62.5	1283	5 Q95RH4	Q95rh4 drosophila
36	85	62.5	1911	5 Q9W4M7	Q9w4m7 drosophila
37	84.5	62.1	233	4 Q8TC53	Q8tc53 homo sapien
38	84	61.8	230	4 Q75263	Q75263 homo sapien
39	84	61.8	326	16 Q91447	Q91447 pseudomonas
40	84	61.8	580	10 Q9M2P6	Q9m2p6 arabidopsis
41	84	61.8	769	5 Q8MRL5	Q8mr15 drosophila
42	84	61.8	1059	5 Q9W493	Q9w493 drosophila
43	84	61.8	1189	4 Q43604	Q43604 homo sapien
44	84	61.8	1199	4 Q60668	Q60668 homo sapien
45	84	61.8	1199	4 Q43487	Q43487 homo sapien

ALIGNMENTS

RESULT 1

Q95PH4 PRELIMINARY; PRT; 1318 AA.
ID Q95PH4;
AC Q95PH4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Histidine kinase Dhkm (Fragment).
GN DHKM.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]_TaxID=44689;
RP SEQUENCE FROM N.A.
RC STRAIN=Ax4;
RA Anjard C., Loomis W.F.;
RT "The histidine kinases of Dictyostelium.";
RL (In) Inouye M., Dutta R. (eds.);
RL HISTIDINE KINASES IN SIGNAL TRANSDUCTION, pp.1-1, Academic press,
RL San Diego (2001).
DR EMBL; AF362374; AAK54093.1;
DR InterPro; IPR003594; ATPbind ATPase.
DR InterPro; IPR005467; His kinase.
DR InterPro; IPR001789; Response_reg.
DR InterPro; IPR000897; SRP54.
DR Pfam; PF02518; HATPase_c; 1.
DR Pfam; PF00072; response_reg; 1.
DR ProDom; PD000039; Response_reg; 1.
DR SMART; SM00387; HATPase_c; 1.
DR SMART; SM00448; REC; 2.
DR PROSITE; PSS0109; HIS_KIN; 1.
DR PROSITE; PSS0110; RESPONSE_REGULATORY; 2.
DR PROSITE; PSS0300; SRP54; 1.
KW Kinase; phosphorylation; Sensory transduction.
FT NON_TER 1
SQ SEQUENCE 1318 AA; 144944 MW; BLDASCA04270666P6 CRC64;

Query Match 72.1%; Score 98; DB 5; Length 1318;
Best Local Similarity 75.0%; Pred. No. 0.00032;

Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 HHKHHKHHKHHKHHH 18
RL Nature 402:761-768 (1999).
RN SEQUENCE FROM N.A.
RP STRAIN=cv. Columbia;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006282; AAD20147.1; -;
DR InterPro; IPR000070; Pectinesterase.
DR Pfam; PF01095; Pectinesterase; 1.
DR PROSITE; PS00503; PECTINESTERASE_2; 1.
SQ SEQUENCE 407 AA; 44973 MW; DEAB7B8E166D0B42 CRC64;

Query Match 70.6%; Score 96; DB 10; Length 407;
Best Local Similarity 68.4%; Pred. No. 0.00019;
Matches 13; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 KKHKKHHKHHKHHKHHK 19
Db 51 KHHHHHHHHHHHHHHHHH 69

RESULT 4

Q8UVJ4 PRELIMINARY; PRT; 404 AA.
AC Q8UVJ4; (Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hoxlw (Fragment).
GN HOKIW.
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
OC NCBI_TaxID=7757;
OX [1]
RN SEQUENCE FROM N.A.
RA Force A., Amores A., Postlethwait J.;
RT "Hox cluster organization in the jawless vertebrate, Petromyzon
marinus, and the evolution of the vertebrate Hox clusters."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; AF434665; AAL61641.1; -;
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
DR DNA-binding; Homeobox; Nuclear protein.
FT NON TER 1 1
FT NON TER 404 404
SQ SEQUENCE 404 AA; 41691 MW; A01537E167772552 CRC64;

Query Match 69.1%; Score 94; DB 13; Length 404;
Best Local Similarity 66.7%; Pred. No. 0.00033;
Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 HHKHHKHHKHHKHHKHHK 20
Db 72 HHHHHHHHHHHHHHHHH 89

RESULT 5

Q26056 PRELIMINARY; PRT; 140 AA.
ID Q26056
AC Q26056;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Histidine-rich protein (Fragment).
OS Plasmodium lophurae.

Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 HHKHHKHHKHHKHHH 18
RL Nature 402:761-768 (1999).
RN SEQUENCE FROM N.A.
RP STRAIN=cv. Columbia;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006282; AAD20147.1; -;
DR InterPro; IPR000070; Pectinesterase.
DR Pfam; PF01095; Pectinesterase; 1.
DR PROSITE; PS00503; PECTINESTERASE_2; 1.
SQ SEQUENCE 407 AA; 44973 MW; DEAB7B8E166D0B42 CRC64;

Query Match 70.6%; Score 96; DB 10; Length 407;
Best Local Similarity 68.4%; Pred. No. 0.00019;
Matches 13; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 KKHKKHHKHHKHHKHHK 19
Db 51 KHHHHHHHHHHHHHHHHH 69

RESULT 4

Q8UVJ4 PRELIMINARY; PRT; 404 AA.
AC Q8UVJ4; (Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hoxlw (Fragment).
GN HOKIW.
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
OC NCBI_TaxID=7757;
OX [1]
RN SEQUENCE FROM N.A.
RA Force A., Amores A., Postlethwait J.;
RT "Hox cluster organization in the jawless vertebrate, Petromyzon
marinus, and the evolution of the vertebrate Hox clusters."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; AF434665; AAL61641.1; -;
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
DR DNA-binding; Homeobox; Nuclear protein.
FT NON TER 1 1
FT NON TER 404 404
SQ SEQUENCE 404 AA; 41691 MW; A01537E167772552 CRC64;

Query Match 69.1%; Score 94; DB 13; Length 404;
Best Local Similarity 66.7%; Pred. No. 0.00033;
Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 HHKHHKHHKHHKHHKHHK 20
Db 72 HHHHHHHHHHHHHHHHH 89

RESULT 5

Q26056 PRELIMINARY; PRT; 140 AA.
ID Q26056
AC Q26056;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Histidine-rich protein (Fragment).
OS Plasmodium lophurae.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 RN NCB1_TaxID=5853;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86174893; PubMed=3007981;
 RA Irving D.O., Cross G.A.M., Feder R., Wallach M.;
 RT "Structure and organization of the histidine-rich protein gene of
 RL Plasmodium lophurae";
 DR Mol. Biochem. Parasitol. 18:223-234(1986).
 RL EMBL; M15317; AAA29616.1; --
 FT NON_TER 1 1
 SQ SEQUENCE 140 AA; 18216 MW; ED05EED8C805B2FA CRC64;
 Query Match 67.6%; Score 92; DB 5; Length 140;
 Best Local Similarity 75.0%; Pred. NO. 0.00022;
 Matches 12; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 3 HHHKHHKHHKHHKHHH 18
 DB 10 HHHHHHHHHHHHHHHH 25
 RESULT 6
 Q8MP30 PRELIMINARY; PRT; 233 AA.
 AC Q8MP30;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
 RN NCB1_TaxID=44689;
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX4;
 RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
 RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
 RA Tungal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
 RT "Sequence and Analysis of Chromosome 2 of Dictyostelium";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC123513; AAM44363.1; --
 KW Hypothetical protein.
 SQ SEQUENCE 233 AA; 28937 MW; AD4C7A2P86BB14B9 CRC64;
 Query Match 67.6%; Score 92; DB 5; Length 233;
 Best Local Similarity 75.0%; Pred. NO. 0.00035;
 Matches 12; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 3 HHHKHHKHHKHHKHHH 18
 DB 71 HHHHHHHHHHHHHHHH 86
 RESULT 7
 Q8SXD4 PRELIMINARY; PRT; 1122 AA.
 ID Q8SXD4; Q9W2R7;
 AC Q8SXD4;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE GH03394p (CG32676 protein).
 GN CG32676 OR CG9725 OR CG9732.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN NCB1_TaxID=7227;
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,

RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Celniker S.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 RN SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA Sutton G.G., Wortman J.R., Richards S., Ashburner M., Henderson S.N.,
 RA Ruan K.H., Doyle C., Baxter E.G., Helt J.G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hosin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milehina N.V., Mobarry C., Morris J., Moehref A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 RN SEQUENCE FROM N.A.
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banron J., An H., Baldwin D., Banron J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreria S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moehref A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of Drosophila melanogaster genome";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN SEQUENCE FROM N.A.
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;

Wed Jan 21 11:27:55 2004

us-10-018-103a-6.rspt

RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY094669; AAM11022.1; -
DR EMBL; AE003451; AAF46623.2; -
DR FlyBase; FBgn0052676; CG32676.
DR InterPro; IPR000626; Ubiquitin.
DR Pfam; PF00240; ubiquitin; 1.
DR SMART; SMO0213; UBO; 1.
DR PROSITE; PS00053; UBIQUITIN 2; 1.
SQ SEQUENCE 1122 AA; 112183 MW; A8FF8B0F79B09B23 CRC64;

Query Match 67.6%; Score 92; DB 5; Length 1122;
Best Local Similarity 75.0%; Pred. No. 0.0014;
Matches 12; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 HHHKHHKHHKHHKHH 18
Db 336 HHHHHHHHHHHHHHHH 351

RESULT 8
ID Q9D6B9 PRELIMINARY; PRT; 109 AA.
AC Q9D6B9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE 3830408D24Rik protein.
GN 3830408D24Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Placenta, and Extraembryonic tissue;
RC MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK014425; BAB29342.1; -
DR MGD; MGI:1917945; 3830408D24Rik.
SQ SEQUENCE 109 AA; 12910 MW; 01AF85E03F763BB0 CRC64;

Query Match 66.9%; Score 91; DB 11; Length 109;
Best Local Similarity 68.4%; Pred. No. 0.00023;
Matches 13; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

Qy 3 HHHKHHKHHKHHKHHK 19
Db 82 HHHHHSPHRLHHKHHK 100

RESULT 9
ID Q9ZW23 PRELIMINARY; PRT; 385 AA.
AC Q9ZW23;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Putative zinc transporter.
GN AT2G29410.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buel C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., Vanaken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Niernan W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.,
RA "Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana.";
RT thaliana.";
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA EMBL; AC004561; AAC95197.1; -
DR InterPro; IPR002524; Cation efflux.
DR Pfam; PF01545; Cation efflux; 1.
DR TIGRFAMs; TIGR01297; CDF; 1.
SQ SEQUENCE 385 AA; 43534 MW; 49A7B564F57FDC2B CRC64;

Query Match 66.9%; Score 91; DB 10; Length 385;
Best Local Similarity 66.7%; Pred. No. 0.00072;
Matches 12; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 3 HHHKHHKHHKHHKHHK 20
Db 201 HHHHHHHKHHKHHK 218

RESULT 10
ID Q813J9 PRELIMINARY; PRT; 4524 AA.
AC Q813J9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF01325W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Devlin K., Baker S., Davies P., Mungal K., Berriman M., Pain A.,
RA Hall N., Bowman S., Churcher C., Quail M., Barrell B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255708; PubMed=12368867;

RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
 RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
 RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
 RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
 RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
 RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
 RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
 RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
 RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
 RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
 RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitz E.,
 RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
 RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
 RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
 RA Sulston J.E., Craig A., Newbold C., Barrell B.G.
 RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."
 RL Nature 419:527-531(2002).
 DR EMBL; AL929354; CAD51629.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 4524 AA; 525287 MW; 542206173C29567A CRC64;

Query Match 65.4%; Score 89.5; DB 5; Length 4524;
 Best Local Similarity 63.6%; Pred. No. 0.0098;
 Matches 14; Conservative 4; Mismatches 1; Indels 3; Gaps 1;

Qy 2 KHHKHK---HHKHHKHHKHK 20
 |||||:|||||:|||||:
 Db 4413 KHHKHKHKHHKHHKHHK 4434

RESULT 11
 Q9BSNO PRELIMINARY; PRT; 325 AA.
 ID Q9BSNO
 AC Q9BSNO
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Similar to RIKEN cdna 2210403L10 gene (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Uterus;
 RA Straussberg R.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC004940; AA04940.1; -.
 FT NON TER 1
 SQ SEQUENCE 325 AA; 36283 MW; 61A24B05707F82FD CRC64;

Query Match 65.4%; Score 89; DB 4; Length 325;
 Best Local Similarity 61.1%; Pred. No. 0.0011;
 Matches 11; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 KHHKHKHHKHHKHHKHH 18
 :|||:|||||:
 Db 303 QRHEHHHHHHHHHHHH 320

RESULT 12
 Q8CAJ8 PRELIMINARY; PRT; 449 AA.
 ID Q8CAJ8
 AC Q8CAJ8
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Naked cuticle 1 homolog.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Head;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 DR EMBL; AK081904; BAC38367.1; -.
 SQ SEQUENCE 449 AA; 50230 MW; FC6ADPBE8836D671 CRC64;

Query Match 65.4%; Score 89; DB 11; Length 449;
 Best Local Similarity 61.1%; Pred. No. 0.0014;
 Matches 11; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 KHHKHKHHKHHKHHKHH 18
 :|||:|||||:
 Db 427 QRHEHHHHHHHHHHHH 444

RESULT 13
 Q969F2 PRELIMINARY; PRT; 451 AA.
 ID Q969F2
 AC Q969F2
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Dvl-binding protein NKD2 (Naked cuticle-2).
 GN NKD2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fetal;
 RX MEDLINE=21490203; PubMed=11604995;
 RA Katoh M.;
 RT "Molecular cloning, gene structure, and expression analyses of NKD1
 RT and NKD2."
 RL Int. J. Oncol. 19:963-969(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21255530; PubMed=11356022;
 RA Wharton K.A. Jr., Zimmermann G., Roussel R., Scott M.P.;
 RT "vertebrate proteins related to Drosophila naked cuticle bind
 RT dishevelled and antagonize wnt signaling."
 RL Dev. Biol. 234:93-106(2001).
 DR EMBL; AB062887; BAB70501.1; -.
 DR EMBL; AF358137; AAK57486.1; -.
 DR Genew; HGNC:17046; NKD2.
 SQ SEQUENCE 451 AA; 50055 MW; F27E708F9FAD1F2A CRC64;

Query Match 65.4%; Score 89; DB 4; Length 451;
 Best Local Similarity 61.1%; Pred. No. 0.0014;
 Matches 11; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 KHHKHKHHKHHKHHKHH 18
 :|||:|||||:
 Db 429 QRHEHHHHHHHHHHHH 446

RESULT 14
 Q9D7U9 PRELIMINARY; PRT; 461 AA.
 ID Q9D7U9
 AC Q9D7U9
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE 2210403L10Rik protein.
 GN NKD2 OR 2210403L10RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Wed Jan 21 11:27:55 2004

Oy 1 KGHKKKKKKKKKKKKKKKK 18
Db 439 QRHHHHHHHHHHHHHHH 456
Search completed: January 20, 2004, 18:26:25
Job time : 25.6377 secs

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Stomach;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamada I.,
RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schrim L.M., Stauber P., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.,
RI "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL; AK008814; BAB25908.1; -.
DR MGD; MGI:1919543; Nkd2.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; ehand; 1.
DR PROSITE; PS00018; EF HAND; 1.
SQ SEQUENCE 461 AA; 51474 MW; 5830BDDA14E868FD CRC64;

Query Match 65.4%; Score 89; DB 11; Length 461;
Best Local Similarity 61.1%; Pred. No. 0.0015;
Matches 11; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Oy 1 KGHKKKKKKKKKKKKKKKK 18
Db 439 QRHHHHHHHHHHHHHHH 456

RESULT 15
Q91Y45 PRELIMINARY; PRT; 461 AA.
ID Q91Y45;
AC Q91Y45;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Naked cuticle-2.
GN Nkd2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6xCBA; TISSUE=Lung;
RX MEDLINE=21255530; PubMed=11356022;
RA Wharton K.A. Jr., Zimmermann G., Rousset R., Scott M.P.;
RT "Vertebrate proteins related to Drosophila naked cuticle bind
RT dishevelled and antagonize wnt signaling."
RL Dev. Biol. 234:93-106(2001).
DR EMBL; AF358136; AAK57485.1; -.
DR MGD; MGI:1919543; Nkd2.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; ehand; 1.
DR PROSITE; PS00018; EF HAND; 1.
SQ SEQUENCE 461 AA; 51501 MW; 7C3B550024E87D27 CRC64;

Query Match 65.4%; Score 89; DB 11; Length 461;
Best Local Similarity 61.1%; Pred. No. 0.0015;
Matches 11; Conservative 3; Mismatches 4; Indels 0; Gaps 0;